Attorney Docket No. 056297-5012-01

PATENT APPLICATION

MYCOBACTERIAL *RpoB* SEQUENCES

Inventor: Thomas Gingeras

Jorg Drenkow

Assignee: AFFYMETRIX, INC.

3380 Central Expressway

Santa Clara, California 95051

a Corporation of California

Entity: Large

PATENT

Attorney Docket No. 056297-5012-01

MYCOBACTERIAL *RpoB* SEQUENCES

STATEMENT OF GOVERNMENT INTEREST

[0001] The work described in this application was supported in part by grant number 1R43a140400 by the NIAID. The Government may have certain rights in this invention.

CROSS-REFERENCE TO RELATED APPLICATIONS

[0002] This application derives priority from USSN 60/080,616, filed April 3, 1998, and incorporated by reference. Applications USSN 08/797,812, filed February 7, 1997, now US Patent 6,228,575; USSN 60/011,339, filed Feb. February 8, 1996; USSN 60/012,631, filed March 1 1996; USSN 08/629,031, filed April 8, 1996, now abandoned; and 60/017,765, filed 15 May 15, 1996 are directed to related subject matter. These applications are specifically incorporated by reference in their entirety for all purposes.

BACKGROUND OF THE INVENTION

Field of the Invention

[0003] This invention is directed to polymorphisms in rpoB genes of mycobacteria and use of the same in the identification and characterization of microorganisms.

Background of the Invention

[0004] Multidrug resistance and human immunodeficiency virus (HIV-1) infections are factors which have had a profound impact on the tuberculosis problem. An increase in the frequency of *Mycobacterium tuberculosis* strains resistant to one or more anti-mycobacterial agents has been reported, Block, et al., (1994) JAMA **271**:665-671. Immunocompromised HIV-1 infected patients not infected with *M. tuberculosis* are frequently infected with *M. avium* complex (MAC) or *M. avium-M. intracellulare* (MAI) complex. These mycobacteria species are often resistant to

the drugs used to treat *M. tuberculosis*. These factors have re-emphasized the importance for the accurate determination of drug sensitivities and mycobacteria species identification.

[0005] In HIV-1 infected patients, the correct diagnosis of the mycobacterial disease is essential since treatment of *M. tuberculosis* infections differs from that called for by other mycobacteria infections, Hoffner, S.E. (1994) *Eur. J. Clin. Microbiol. Inf. Dis.* **13**:937-941. Non-tuberculosis mycobacteria commonly associated with HIV-1 infections include *M. kansasii*, *M. xenopi*, *M. fortuitum*, *M. avium* and *M. intracellulare,* Wolinsky, E., (1992) *Clin. Infect. Dis.* **15**:1-12, Shafer, R.W. and Sierra, M.F. 1992 *Clin. Infect. Dis.* **15**:161-162. Additionally, 13% of new cases (HIV-1 infected and non-infected) of *M. tuberculosis* are resistant to one of the primary anti-tuberculosis drugs (isoniazid [INH], rifampin [RIF], streptomycin [STR], ethambutol [EMB] and pyrazinamide [PZA] and 3.2% are resistant to both RIF and INH, Block, et al., *JAMA* **271:**665-671, (1994). Consequently, mycobacterial species identification and the determination of drug resistance have become central concerns during the diagnosis of mycobacterial diseases.

[0006] Methods used to detect, and to identify *Mycobacterium* species vary considerably. For detection of *Mycobacterium tuberculosis*, microscopic examination of acid-fast stained smears and cultures are still the methods of choice in most microbiological clinical laboratories. However, culture of clinical samples is hampered by the slow growth of mycobacteria. A mean time of four weeks is required before sufficient growth is obtained to enable detection and possible identification. Recently, two more rapid methods for culture have been developed involving a radiometric, Stager, C.E. et al., (1991) *J. Clin. Microbiol.* **29**:154-157, and a biphasic (broth/agar) system Sewell, et al., (1993) *J. Clin. Microbiol.* **29**:2689-2472. Once grown, cultured mycobacteria can be analyzed by lipid composition, the use of species specific antibodies, species specific DNA or RNA probes and PCR-based sequence analysis of 16S rRNA gene (Schirm, et al. (1995) *J. Clin. Microbiol.* **33**:3221-3224; Kox, et al. (1995) *J. Clin. Microbiol.* **33**:3225-3233) and IS6110 specific repetitive sequence analysis (For a review see, e.g., Small et al., P.M. and van Embden, J.D.A. (1994) *Am. Society for Microbiology*, pp. 569-582). The analysis of 16S rRNA sequences (RNA and DNA) has been the most informative molecular approach to identify *Mycobacteria* species (Jonas, et al., *J. Clin. Microbiol.* **31**:2410-

2416 (1993)). However, to obtain drug sensitivity information for the same isolate, additional protocols (culture) or alternative gene analysis is necessary.

[0007] To determine drug sensitivity information, culture methods are still the protocols of choice. *Mycobacteria* are judged to be resistant to particular drugs by use of either the standard proportional plate method or minimal inhibitory concentration (MIC) method. However, given the inherent lengthy times required by culture methods, approaches to determine drug sensitivity based on molecular genetics have been recently developed.

[0008] Because resistance to RIF in *E. coli* strains was observed to arise as a result of mutations in the *rpo*B gene, Telenti, et al., id., identified a 69 base pair (bp) region of the *M. tuberculosis rpo*B gene as the locus where RIF resistant mutations were focused. Kapur, et al., (1995) *Arch. Pathol. Lab. Med.* **119**:131-138, identified additional novel mutations in the *M. tuberculosis rpo*B gene which extended this core region to 81 bp. In a detailed review on antimicrobial agent resistance in mycobacteria, Musser (*Clin. Microbiol. Rev.*, **8**:496-514 (1995)), summarized all the characterized mutations and their relative frequency of occurrence in this 81 bp region of *rpo*B. Missense mutations comprise 88% of all known mutations while insertions (3 or 6 bp) and deletions (3, 6 and 9 bp) account for 4% and 8% of the remaining mutations, respectively. Approximately 90% of all RIF resistant tuberculosis isolates have been shown to have mutations in this 81 bp region. The remaining 10% are thought possibly to involve genes other than *rpo*B.

[0009] For the above reasons, it would be desirable to have simpler methods which identify and characterize microorganisms, such as *Mycobacteria,* both at the phenotypic and genotypic level. This invention fulfills that and related needs.

SUMMARY OF THE INVENTION

[0010] In one aspect, the invention provides isolated nucleic acids comprising at least 25, 50, 75, 100, or 200 contiguous bases from an rpoB sequence shown in Table 1 (SEQ ID NOS: 1-181). Some nucleic acid comprise a complete sequence shown in Table 1.

[0011] The invention further provides a set of probes perfectly complementary to and spanning such nucleic acids, preferably spanning one of the complete sequences shown in Table 1 (SEQ ID NOS: 1-181).

[0012] The invention further provides methods of classifying mycobacteria. Some such methods entail providing a sample comprising a mycobacterial rpoB target nucleic acid from a mycobacteria, determining the sequence of a segment of at least 50 contiguous bases from the target nucleic acid; comparing the determined sequence to at least one sequence shown in Table 1; and classifying the mycobacteria from the extent of similarity of the compared sequences. Preferably, at least 100 or 200 contiguous bases are determined from the target nucleic acid. Preferably, the determined sequence is compared with a plurality of sequences from Table 1, for example, 10, 20, 50 or all of the sequence from Table 1 (SEQ ID NOS: 1-181).

[0013] In other methods of classification, the identity of one or more bases in the target sequence at one or more positions corresponding to one or more of the highlighted positions in a sequence shown in Table 1 is determined. The identity of the one or more bases characterizing the species of mycobacteria that is present in the sample. In some methods, the identity of at least 10 bases in the target nucleic acid at positions corresponding to highlighted positions in a sequence shown in Table 1 is determined. In some methods, the identity of at least 20 bases in the target sequence at highlighted positions shown in Table 1 are identified. In some methods, at least 20 determined bases are compared with 20 bases occupying corresponding positions in each of at least ten sequences from Table 1.

[0014] In another aspect, the invention provides sequence-specific polynucleotide probes or primers that hybridizes to a segment of a mycobacterial rpoB sequence shown in Table 1 or its complement without hybridizing to the M. tuberculosis sequence designated ATCC9-Mtb in Table 1 or its complement, the segment including a highlighted nucleotide position shown in Table 1. In some such probes, a central position of the probe aligns with a highlighted nucleotide position shown in Table 1. In some such primers, the 3' end of the primer aligns with a highlighted nucleotide position shown in Table 1. Some probes and primers are between 10 and 50 bases long.

[0015] In another aspect, the invention provides a computer-readable storage medium for storing data for access by an application program being executed on a data processing system. Such a system comprises a data structure stored in the computer-readable storage medium. The data structure includes information resident in a database used by the application program and includes a plurality of records, each record comprising information identifying a polymorphism

or sequence shown in Table 1. Some records have a field identifying a base occupying a polymorphic site and a field identifying location of the polymorphic site. Some records record a contiguous segment of at least 50, 100, or 200 bases from an rpoB sequence shown in Table 1. Some storage medium comprise at least ten records each recording a contiguous segment of at least 50 bases from at least ten rpoB sequences shown in Table 1.

BRIEF DESCRIPTION OF THE DRAWINGS

[0016] Fig. 1: Computer that may be utilized to execute software embodiments of the present invention.

[0017] Fig. 2: A system block diagram of a typical computer system that may be used to execute software embodiments of the invention.

DEFINITIONS

[0018] A polynucleotide can be DNA or RNA, and single- or double-stranded. Polynucleotide can be naturally occurring or synthetic, and can be of any length. Preferred polynucleotide probes of the invention include contiguous segments of DNA, or their complements including any of the highlighted bases shown in Table 1. The segments are usually between 5 and 100 bases, and often between 5-10, 5-20, 10-20, 10-50, 20-50 or 20-100 bases. The highlighted site can occur within any position of the segment. Preferred polynucleotide probes are capable of binding in a base-specific manner to a complementary strand of nucleic acid. Such probes include peptide nucleic acids, as described in Nielsen et al., *Science* 254, 1497-1500 (1991), and probes having nonnaturally occurring bases.

[0019] The term primer refers to a single-stranded polynucleotide capable of acting as a point of initiation of template-directed DNA synthesis under appropriate conditions (*i.e.*, in the presence of four different nucleoside triphosphates and an agent for polymerization, such as, DNA or RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. The appropriate length of a primer depends on the intended use of the primer but typically ranges from 15 to 30 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. The term primer site refers to the area of the target DNA to which a primer hybridizes. The term primer pair

means a set of primers including a 5' upstream primer that hybridizes with the 5' end of the DNA sequence to be amplified and a 3', downstream primer that hybridizes with the complement of the 3' end of the sequence to be amplified.

[0020] A cDNA or cRNA is derived from an RNA if it produced by a process in which the RNA serves as a template for production of the cDNA or cRNA.

[0021] Hybridizations are usually performed under stringent conditions, for example, at a salt concentration of no more than 1 M and a temperature of at least 25°C. For example, conditions of 5 x SSPE (750 mM NaCl, 50 mM Na Phosphate, 5 mM EDTA, pH 7.4) and a temperature of 25-30°C are suitable for allele-specific probe hybridizations.

[0022] An isolated nucleic acid means an object species invention that is the predominant species present (*i.e.*, on a molar basis it is more abundant than any other individual species in the composition). Preferably, an isolated nucleic acid comprises at least about 50, 80 or 90 percent (on a molar basis) of all macromolecular species present. Most preferably, the object species is purified to essential homogeneity (contaminant species cannot be detected in the composition by conventional detection methods).

[0023] For sequence comparison and homology determination, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

[0024] Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by visual inspection (*see generally*, Ausubel *et al., infra*).

[0025] One example of algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul *et al., J. Mol.*

Biol. 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915).

[0026] In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.*, Karlin & Altschul (1993) *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

[0027] The term "target nucleic acid" refers to a nucleic acid (often derived from a biological sample), to which the probe nucleic acid is designed to specifically hybridize. It is the presence or expression level of the target nucleic acid that is to be detected or quantified. The target nucleic acid has a sequence that is complementary to the nucleic acid sequence of the corresponding probe directed to the target. The term target nucleic acid may refer to the specific subsequence of a larger nucleic acid to which the probe is directed or to the overall sequence (*e.g.* gene or mRNA) whose expression level it is desired to detect. The difference in usage will be apparent from context.

[0028] "Subsequence" refers to a sequence of nucleic acids that comprise a part of a longer sequence of nucleic acids.

DETAILED DESCRIPTION

I. Mycobacterial Sequences of rpoB Genes

[0029] Table 1 shows a comparison of a substantial collection of mycobacterial strains of an about 700-nucleotide conserved region of an rpoB gene. The sequences shown in Table 1 are identified as follows: SEQ ID NOS: 1-56, respectively, are shown on pages 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61 and 65; SEQ ID NOS: 57-112, respectively, are shown on pages 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62 and 65; SEQ ID NOS: 113-168, respectively, are shown on pages 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63 and 66; SEQ ID NOS: 169-181, respectively, are shown on pages 24, 28, 32, 36, 40, 44, 52, 56, 60, 64 and 68. The first sequence, designated as a reference sequence, is from *M. tuberculosis*. Nucleotides are numbered consecutively starting from the first nucleotide of the reference sequences. Other sequences are from other strains of mycobacteria. For example, the sequences designated ATCC-av, M29, M30...M104 are from *M. avium*. Sequences designated from ATT-chelnew, M11, M13, and M17 are from *M. chelonae*. Sequences designated ATCC-for, M53, M55, M56, and M74 are from *M. fortuitum*, and so forth. Complete correspondence between strain designations and strain types is shown in Table 2. Nucleotides in a mycobacterial sequence are accorded the same number as the corresponding position of the reference sequence when the two are maximally aligned. Differences between a sequence and the reference sequences are shown in highlighted type. Many of the highlighted positions are common to all tested members of a species. Other highlighted positions vary

among different isolates in a species. Both types of variation can be useful in speciation analysis.

II. Analysis of Species Variations

A. Preparation of Samples

[0030] An rpoB sequence is isolated from a sample of an unknown mycobacteria being tested. Nucleic acids can be isolated from myobacteria by standard methods as described in WO 97/29212 (incorporated by reference in its entirety for all purposes). The rpoB sequences to be analyzed can then be isolated and amplified by means of PCR. *See generally PCR Technology: Principles and Applications for DNA Amplification* (ed. H.A. Erlich, Freeman Press, NY, NY, 1992); *PCR Protocols: A Guide to Methods and Applications* (eds. Innis, et al., Academic Press, San Diego, CA, 1990); Mattila et al., *Nucleic Acids Res.* 19, 4967 (1991); Eckert et al., *PCR Methods and Applications* 1, 17 (1991); *PCR* (eds. McPherson et al., IRL Press, Oxford); and U.S. Patent 4,683,202 (each of which is incorporated by reference for all purposes). Primers for PCR preferably flank the regions of interest rpoB genes, although primers to internal sites can be used if it is intended to analyze only certain sites of potential species variation. Exemplary primers are described in WO 97/29212. If necessary, additional sequences flanking the sequences shown in Table 1 can be determined using probes based on the sequences in Table 1 to isolate full-length rpoB sequences from the appropriate mycobacterial species.

B. Detection of Species-Specific Variations in Target DNA

1. Sequence-Specific Probes

[0031] The design and use of sequence-specific probes for analyzing polymorphisms is described by e.g., Saiki et al., *Nature* 324, 163-166 (1986); Dattagupta, EP 235,726, Saiki, WO 89/11548. Sequence-specific probes can be designed that hybridize to a segment of target DNA in one isolate of mycobacteria that do not isolate to a corresponding isolate in another due to the presence of allelic or species variations in the respective segments from the two sequences. Hybridization conditions should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the sequences. Some probes are designed to hybridize to a

segment of target DNA such that the site of potential sequence variation aligns with a central position (e.g., in a 15 mer at the 7 position; in a 16 mer, at either the 8 or 9 position) of the probe. This design of probe achieves good discrimination in hybridization between different allelic and species variants.

[0032] Sequence-specific probes are often used in pairs, one member of a pair showing a perfect match to a reference form of a target sequence and the other member showing a perfect match to a variant form. Several pairs of probes can then be immobilized on the same support for simultaneous analysis of multiple potential variations within the same target sequence.

2. Tiling Arrays

[0033] The bases occupying sites of potential variation can also be identified by hybridization to nucleic acid arrays, some example of which are described by WO 95/11995 (incorporated by reference in its entirety for all purposes). Such arrays contain a series of overlapping probes spanning a reference sequence. Any of the rpoB sequences shown in Table 1, or contiguous segments of, for example, at least 25, 50, 100 or 200 bases thereof, can serve as a reference sequence. WO 95/11995 also describes subarrays that are optimized for detection of a variant forms of a precharacterized polymorphism. Such a subarray contains probes designed to be complementary to a second reference sequence, which is a variant of the first reference sequence. The inclusion of a second group (or further groups) can be particular useful for analyzing short subsequences of the primary reference sequence in which multiple mutations are expected to occur within a short distance commensurate with the length of the probes (*i.e.*, two or more mutations within 9 to 21 bases).

3. Sequence-Specific Primers

[0034] A sequence-specific primer hybridizes to a site on target DNA overlapping a polymorphism and only primes amplification of a variant form to which the primer exhibits perfect complementarily. See Gibbs, *Nucleic Acid Res.* 17, 2427-2448 (1989). This primer is used in conjunction with a second primer which hybridizes at a distal site. Amplification proceeds from the two primers leading to a detectable product signifying the particular variant form is present. A control is usually performed with a second pair of primers, one of which

shows a single base mismatch at the site of variation and the other of which exhibits perfect complementarily to a distal site. The single-base mismatch prevents amplification and no detectable product is formed. The method works best when the mismatch is included in the 3'-most position of the primer aligned with the point of variation because this position is most destabilizing to elongation from the primer. *See, e.g.,* WO 93/22456.

4. Direct-Sequencing

[0035] The direct analysis of mycobacterial sequences can be accomplished using either the dideoxy chain termination method or the Maxam Gilbert method (see Sambrook et al., *Molecular Cloning, A Laboratory Manual* (2nd Ed., CSHP, New York 1989); Zyskind et al., *Recombinant DNA Laboratory Manual*, (Acad. Press, 1988)).

III. Methods of Use

[0036] The sequences and polymorphisms shown in Table 1 are useful for identifying the presence of myobacteria in samples, and optionally, classifying the mycobacteria. The sample can be obtained from a patient or from a biological source, such as a food product.

[0037] The sequences shown in Table 1 can be used for design of sequence-specific probes or primers encompassing polymorphic sites as described above. These probes or primers can then be used to determine the base occupying a corresponding position in an rpoB sequence from an isolate in a sample under test. A base in one sequence corresponds with a base in another when the two bases occupy the same position when the two sequences are maximally aligned by one of the criteria described in Definitions.

[0038] Alternatively, the sequences shown in Table 1 can be used for design of tiling arrays in which one or more of the sequences serves as a reference sequence. At least one set of overlapping probes is designed spanning a segment of the reference sequence, as described in WO95/11995 or EP 717,113. Target sequences from samples under test can be hybridized to such arrays, optionally in combination with controls of known rpoB sequences. The hybridization pattern of a target sequence to such an array can be analyzed to determine the identity of bases at which the target sequence differs from the reference sequence, as described in WO 95/11995.

[0039] One or more of the above methods, or direct sequencing, can be used to identify the base occupying at least one and usually several (e.g., 5, 10, 15, 25, 50 or 100) sites of potential variation between the 16S RNA and/or rpoB gene in an unknown mycobacteria relative to bases occupying corresponding sites in one or more known strains of mycobacteria, such as those shown in Table 1. This analysis results in a profile of bases occupying particular sites that characterizes the mycobacterial strain under test. The profile is compared with the corresponding profiles of different mycobacterial isolates shown in e.g., Table 1. In general, the unknown mycobacterium isolate is characterized as being from the same mycobacterial species as the precharacterized isolate with which it shares the greatest similarity in base profile.

[0040] In some methods, the sequence of a contiguous segment of the rpoB target nucleic acid is determined in a sample under test for comparison with one or more of the sequences shown in Table 1. The mycobacteria is classified by the extent of similarity. For example, if a target nucleic acid shows greater sequence identity to rpoB sequences from one species than any other, the sample from which the target was obtained is typically classified as arising from that species.

[0041] Alternatively, an array of tiled probes based on a reference sequence shown in Table 1 can be used for identifying and characterizing mycobacterial sequences based on comparison of hybridization patterns. Such an array is hybridized to a 16S RNA or rpoB target sequence from a sample, and the hybridization pattern compared with the hybridization pattern of one or more control sequences. The hybridization patterns of control sequences can be historic controls, stored, for example, in a computer database, or can be contemporaneous controls performed at or near the same time as the hybridization to the target sequence. Optionally, hybridization of target and reference sequence can be performed simultaneously using different labels.

[0042] Method of classifying unknown mycobacterial isolate by matching the hybridization pattern of a target sequence with those of control sequences from characterized species are described in more detail in WO 97/29212 (incorporated by reference in its entirety for all purposes). In an idealized case, the detection of a particular hybridization pattern in an isolate characterizes that isolate as belonging to a particular species. This can occur when the hybridization pattern detected in the isolate is uniquely associated with a specific species. More frequently however, such an unique one-to-one correspondence is not present. Instead, the hybridization pattern observed in an isolate does not bear a unique correspondence with a

previously characterized species. However, the hybridization pattern detected is associated with a probability of the organism being screened belonging to a particular species (or not) or carrying a particular phenotypic trait (or not). As a result, analysis of an increasing number of polymorphic sites in an isolate, allows one to classify the isolated with an increasing level of confidence. Algorithms can be used to derive such composite probabilities from the comparison of multiple polymorphic forms between an isolate and references. Typically, the mathematical algorithm makes a call of the identity of the species and assign a confidence level to that call. One can determine the confidence level (>90%, >95% etc.) that one desires and the algorithm will analyze the hybridization pattern and either provide an identification or not. Occasionally, the call is that the sample may be one of two, three or more species, in which case a specific identification is not be possible. However, one of the strengths of this technique is that the rapid screening made possible by the chip-based hybridization allows one to continuously expand a database of patterns ultimately to enable the identification of species previously unidentifiable due to lack of sufficient information.

IV. Modified Polypeptides and Gene Sequences

[0043] The invention further provides variant forms of nucleic acids and corresponding proteins. The nucleic acids comprise one of the sequences described in Table 1. Some nucleic acid encode full-length variant forms of proteins. Variant proteins have the prototypical amino acid sequences of encoded by nucleic acid sequence shown in Table 1 (read so as to be in-frame with the full-length coding sequence of which it is a component).

[0044] Variant genes can be expressed in an expression vector in which a variant gene is operably linked to a native or other promoter. Usually, the promoter is a eukaryotic promoter for expression in a mammalian cell. The transcription regulation sequences typically include a heterologous promoter and optionally an enhancer which is recognized by the host. The selection of an appropriate promoter, for example trp, lac, phage promoters, glycolytic enzyme promoters and tRNA promoters, depends on the host selected. Commercially available expression vectors can be used. Vectors can include host-recognized replication systems, amplifiable genes, selectable markers, host sequences useful for insertion into the host genome, and the like.

[0045] The means of introducing the expression construct into a host cell varies depending upon the particular construction and the target host. Suitable means include fusion, conjugation, transfection, transduction, electroporation or injection, as described in Sambrook, *supra*. A wide variety of host cells can be employed for expression of the variant gene, both prokaryotic and eukaryotic. Suitable host cells include bacteria such as *E. coli*, yeast, filamentous fungi, insect cells, mammalian cells, typically immortalized, *e.g.*, mouse, CHO, human and monkey cell lines and derivatives thereof. Preferred host cells are able to process the variant gene product to produce an appropriate mature polypeptide. Processing includes glycosylation, ubiquitination, disulfide bond formation, general post-translational modification, and the like.

[0046] The protein may be isolated by conventional means of protein biochemistry and purification to obtain a substantially pure product, *i.e.*, 80, 95 or 99% free of cell component contaminants, as described in Jacoby, *Methods in Enzymology* Volume 104, Academic Press, New York (1984); Scopes, *Protein Purification, Principles and Practice*, 2nd Edition, Springer-Verlag, New York (1987); and Deutscher (ed), *Guide to Protein Purification, Methods in Enzymology,* Vol. 182 (1990). If the protein is secreted, it can be isolated from the supernatant in which the host cell is grown. If not secreted, the protein can be isolated from a lysate of the host cells.

[0047] In addition to substantially full-length polypeptides expressed by variant genes, the present invention includes biologically active fragments of the polypeptides, or analogs thereof, including organic molecules which simulate the interactions of the peptides. Biologically active fragments include any portion of the full-length polypeptide which confers a biological function on the variant gene product, including ligand binding, and antibody binding. Ligand binding includes binding by nucleic acids, proteins or polypeptides, small biologically active molecules, or large cellular structures.

[0048] Polyclonal and/or monoclonal antibodies that specifically bind to variant gene products but not to corresponding prototypical gene products are also provided. Antibodies can be made by injecting mice or other animals with the variant gene product or synthetic peptide fragments thereof. Monoclonal antibodies are screened as are described, for example, in Harlow & Lane, *Antibodies, A Laboratory Manual*, Cold Spring Harbor Press, New York (1988); Goding, *Monoclonal antibodies, Principles and Practice* (2d ed.) Academic Press, New York (1986).

Monoclonal antibodies are tested for specific immunoreactivity with a variant gene product and lack of immunoreactivity to the corresponding prototypical gene product. These antibodies are useful in diagnostic assays for detection of the variant form, or as an active ingredient in a pharmaceutical composition.

V. Kits

[0049] The invention further provides kits comprising at least one sequence-specific probe as described above. Often, the kits contain one or more pairs of sequence-specific probes hybridizing to different forms of a polymorphism. In some kits, the sequence-specific probes are provided immobilized to a substrate. For example, the same substrate can comprise sequence-specific probes for detecting at least 10, 100 or all of the variations shown in Table 1. Optional additional components of the kit include, for example, restriction enzymes, reverse-transcriptase or polymerase, the substrate nucleoside triphosphates, means used to label (for example, an avidin-enzyme conjugate and enzyme substrate and chromogen if the label is biotin), and the appropriate buffers for reverse transcription, PCR, or hybridization reactions. Usually, the kit also contains instructions for carrying out the methods.

VI. Computer Databases

[0050] Fig. 1 illustrates an example of a computer system that can be used to store records relating to polymorphisms of the invention and perform algorithms comparing polymorphic profiles and to classify species. Fig. 2 shows a computer system 100 which includes a monitor 102, screen 104, cabinet 106, keyboard 108, and mouse 110. Mouse 110 may have one or more buttons such as mouse buttons 112. Cabinet 106 houses a CD-ROM drive 114, a system memory and a hard drive (see Fig. 2) which can be utilized to store and retrieve software programs incorporating code that implements the present invention, data for use with the present invention, and the like. Although a CD-ROM 116 is shown as an exemplary computer readable storage medium, other computer readable storage media including floppy disks, tape, flash memory, system memory, and hard drives may be utilized. Cabinet 106 also houses familiar computer components such as a central processor, system memory, hard disk, and the like.

[0051] Fig. 2 shows a system block diagram of computer system 100 that may be used to execute software embodiments of the present invention. As in Fig. 1, computer system 100 includes monitor 102 and keyboard 108. Computer system 100 further includes subsystems such as a central processor 102, system memory 120, I/O controller 122, display adapter 124, removable disk 126 (e.g., CD-ROM drive), fixed disk 128 (e.g., hard drive), network interface 130, and speaker 132. Other computer systems suitable for use with the present invention may include additional or fewer subsystems. For example, another computer system can include more than one processor 102 (i.e., a multi-processor system) or a cache memory.

[0052] Arrows such as 134 represent the system bus architecture of computer system 100. However, these arrows are illustrative of any interconnection scheme serving to link the subsystems. For example, a local bus can be utilized to connect the central processor to the system memory and display adapter. Computer system 100 shown in Fig. 1 is but an example of a computer system suitable for use with the present invention.

[0053] The computer stores records relating to the polymorphisms of the record. Some such records record a polymorphism by reference to the position of a polymorphic site and the identity of base(s) occupying that site in one or more species. Some databases include records for at least ten polymorphic sites in at least ten of the sequences shown in Table 1. Some databases include records for all of the polymorphic sites in at least one of the sequences shown in Table 1. Some databases includes records for at least 100, 1000, or 2000 polymorphic sites shown in Table 1. Some databases include records for all of the polymorphic sites shown in Table 1.

[0054] The foregoing invention has been described in some detail by way of illustration and example, for purposes of clarity and understanding. It will be obvious to one of skill in the art that changes and modifications may be practiced within the scope of the appended claims. Therefore, it is to be understood that the above description is intended to be illustrative and not restrictive. The scope of the invention should, therefore, be determined not with reference to the above description, but should instead be determined with reference to the following appended claims, along with the full scope of equivalents to which such claims are entitled.

[0055] All patents, patent applications and publications cited in this application are hereby incorporated by reference in their entirety for all purposes to the same extent as if each individual patent, patent application or publication were so individually denoted.

Table 1

BASE NOS 1-60

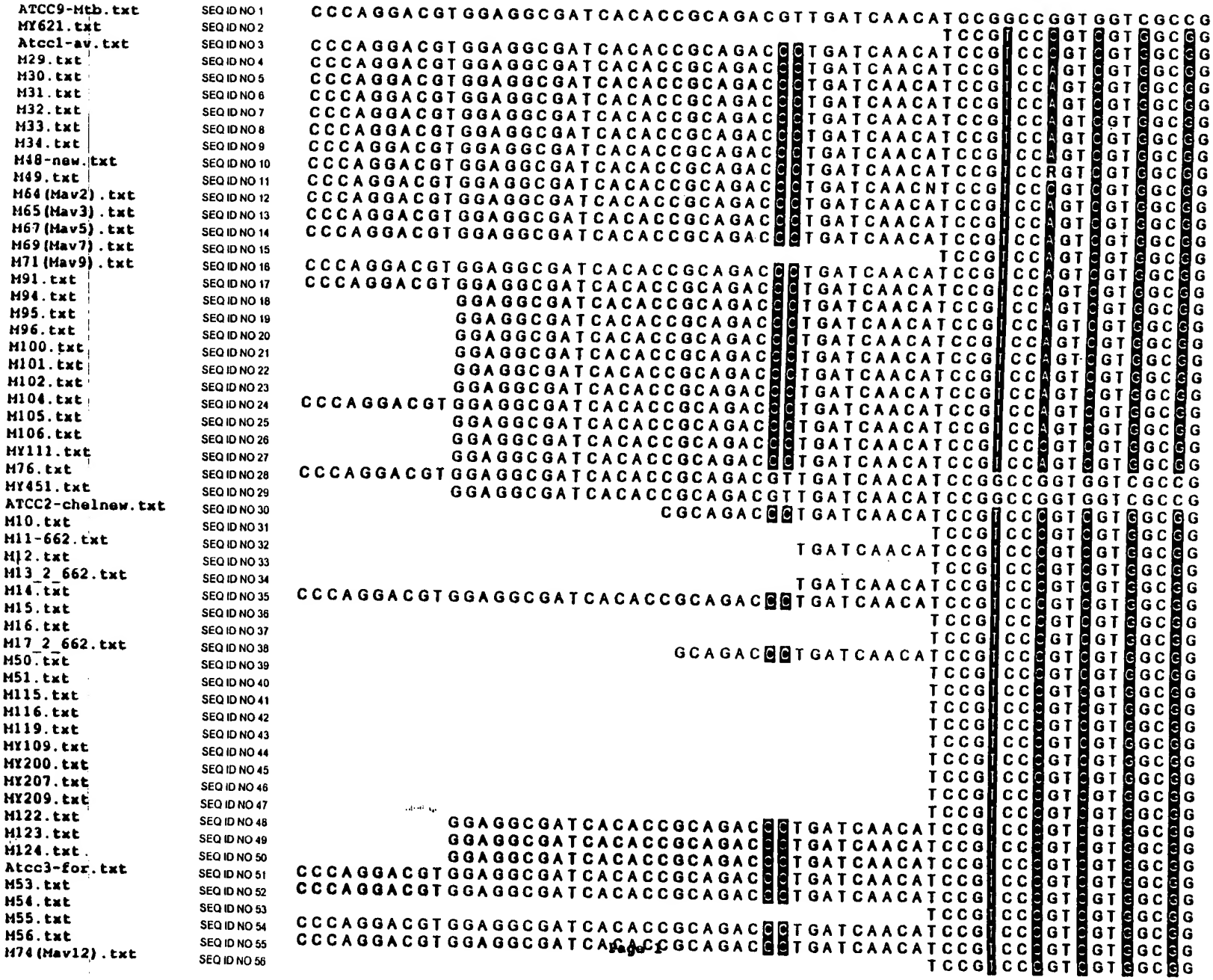

Table 1

BASE NOS 1-60

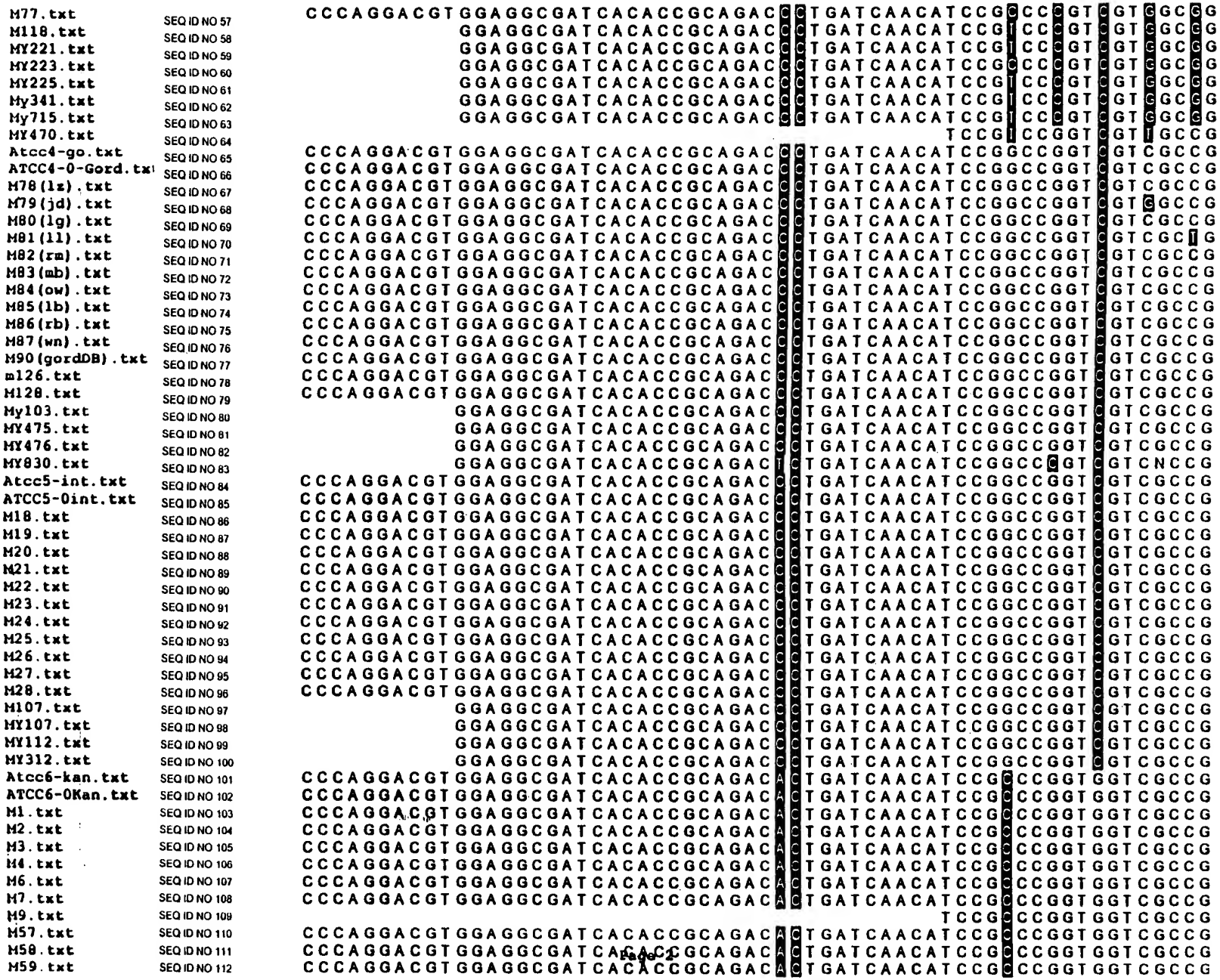

Table 1

BASE NOS 1-60

M60.txt	SEQ ID NO 113	CCCAGGACGTGGAGGCGATCACACCGCAGACACTGATCAACATCCGGCCGGTGGTCGCCG
M61.txt	SEQ ID NO 114	CCCAGGACGTGGAGGCGATCACACCGCAGACACTGATCAACATCCGGCCGGTGGTCGCCG
M62.txt	SEQ ID NO 115	CCCAGGACGTGGAGGCGATCACACCGCAGACACTGATCAACATCCGGCCGGTGGTCGCCG
MY106.txt	SEQ ID NO 116	GGAGGCGATCACACCGCAGACACTGATCAACATCCGGCCGGTGGTCGCCG
MY216.txt	SEQ ID NO 117	GGAGGCGATCACACCGCAGACACTGATCAACATCCGGCCGGTGGTCGCCG
MY218.txt	SEQ ID NO 118	GGAGGCGATCACACCGCAGACACTGATCAACATCCGGCCGGTGGTCGCCG
MY226.txt	SEQ ID NO 119	GGAGGCGATCACACCGCAGACACTGATCAACATCCGGCCGGTGGTCGCCG
M109.txt	SEQ ID NO 120	GGAGGCGATCACACCGCAGACGCTGATCAACATCCGGCCGGTGGTCGCCG
M111.txt	SEQ ID NO 121	GGAGGCGATCACACCGCAGACGCTGATCAACATCCGGCCGGTGGTCGCCG
M112.txt	SEQ ID NO 122	GGAGGCGATCACACCGCAGACGCTGATCAACATCCGGCCGGTGGTCGCCG
M113.txt	SEQ ID NO 123	GGAGGCGATCACACCGCAGACGCTGATCAACATCCGGCCGGTGGTCGCCG
M114.txt	SEQ ID NO 124	GGAGGCGATCACACCGCAGACGCTGATCAACATCCGGCCGGTGGTCGCCG
MY325.txt	SEQ ID NO 125	GGAGGCGATCACACCGCAGACGCTGATCAACATCCGGCCGGTGGTCGCCG
MY718.txt	SEQ ID NO 126	GGAGGCGATCACACCGCAGACGCTGATCAACATCCGGCCGGTGGTCGCCG
MY214.txt	SEQ ID NO 127	GGAGGCGATCACACCGCAGACGTTGATCAACATCCGTCCGGTCGTTGCCG
MY224.txt	SEQ ID NO 128	GGAGGCGATCACACCGCAGACGTTGATCAACATCCGTCCGGTCGTTGCCG
My244.txt	SEQ ID NO 129	GGAGGCGATCACACCGCAGACGCTGATCAACATCCGTCCGGTCGTTGCCG
My339.txt	SEQ ID NO 130	GGAGGCGATCACACCGCAGACGCTGATCAACATCCGTCCGGTCGTTGCCG
My343.txt	SEQ ID NO 131	GGAGGCGATCACACCGCAGACGCTGATCAACATCSGTCCGGTCGTTGCCG
MY458.txt	SEQ ID NO 132	TCCGTCCCGTCGTGGCGG
MY809.txt	SEQ ID NO 133	TCCGTCCCGTCGTGGCGG
MY817.txt	SEQ ID NO 134	TCCGTCCCGTCGTGGCGG
MY821.txt	SEQ ID NO 135	TCCGTCCNGTCGTGGCGG
MY824.txt	SEQ ID NO 136	TCCGTCCCGTCGTGGCGG
MY102.txt	SEQ ID NO 137	GGAGGCGATCACACCGCAGACCCTGATCAACATCCGCCCGGTGGTCGCCG
My105.txt	SEQ ID NO 138	GGAGGCGATCACACCGCAGACCCTGATCAACATCCGCCCGGTGGTCGCCG
MY251.txt	SEQ ID NO 139	GGAGGCGATCACACCGCAGACCCTGATCAACATCCGTCCGGTGGTCGCCG
My256.txt	SEQ ID NO 140	GGAGGCGATCACACCGCAGACCCTGATCAACATCCGTCCGGTGGTCGCCG
My294.txt	SEQ ID NO 141	GGAGGCGATCACACCGCAGACCCTGATCAACATCCGCCCGGTGGTCGCCG
MY354.txt	SEQ ID NO 142	GGAGGCGATCACACCGCAGACCCTGATCAACATCCGCCCGGTGGTCGCCG
Atcc7-scr.txt	SEQ ID NO 143	CCCAGGACGTGGAGGCGATCACACCGCAGACCCTGATCAACATCCGGCCGGTCGTGGCCG
ATCC7-0scr.txt	SEQ ID NO 144	CCCAGGACGTGGAGGCGATCACACCGCAGACCCTGATCAACATCCGGCCGGTCGTGGCCG
MY121.txt	SEQ ID NO 145	GGAGGCGATCACACCGCAGACCCTGATCAACATCCGGCCGGTGGTGGCCG
MY249.txt	SEQ ID NO 146	GGAGGCGATCACACCGCAGACCCTGATCAACATCCGGCCGGTCGTGGCCG
MY372.txt	SEQ ID NO 147	GGAGGCGATCACACCGCAGACGCTGATCAACATCCGGCCGGTCGTCGCCG
MY378.txt	SEQ ID NO 148	GGAGGCGATCACACCGCAGACCCTGATCAACATCCGGCCGGTCGTGGCCG
MY484.txt	SEQ ID NO 149	GGAGGCGATCACACCGCAGACCCTGATCAACATCCGTCCGGTCGTGGCGG
My556.txt	SEQ ID NO 150	GGAGGCGATCACACCGCAGACCCTGATCAACATCCGTCCGGTCGTGGCGG
MY563.txt	SEQ ID NO 151	GGAGGCGATCACACCGCAGACCCTGATCAACATCCGTCCGGTCGTGGCGG
MY586.txt	SEQ ID NO 152	GGAGGCGATCACACCGCAGACCCTGATCAACATCCGTCCGGTCGTGGCGG
Atcc8-sme.txt	SEQ ID NO 153	CCCAGGACGTGGAGGCGATCACACCGCAGACCCTGATCAACATCCGTCCCGTCGTGGCGG
M35.txt	SEQ ID NO 154	TCCGTCCGNTCGTGGCGG
M36.txt	SEQ ID NO 155	CCCAGGACGTGGAGGCGATCACACCGCAGACCCTGATCAACATCCGCCCCGTCGTGGCGG
M37.txt	SEQ ID NO 156	CCCAGGACGTGGAGGCGATCACACCGCAGACCCTGATCAACATCCGTCCCGTCGTGGCGG
m125.txt	SEQ ID NO 157	CCCAGGACGTGGAGGCGATCACACCGCAGACCCTGATCAACATCCGGCCCGTCGTCGCCG
MY143.txt	SEQ ID NO 158	GGAGGCGATCACACCGCAGACCCTGATCAACATCCGTCCCGTCGTGGCGG
My104.txt	SEQ ID NO 159	GGAGGCGATCACACCGCAGACCCTGATCAACATCCGGCCCGTCGTCGCCG
MY196.txt	SEQ ID NO 160	GGAGGCGATCACACCGCAGACCCTGATCAACATCCGGCCCGTCGTCGCCG
MY357.txt	SEQ ID NO 161	GGAGGCGATCACACCGCAGACCCTGATCAACATCCGGCCCGTCGTCGCCG
My358.txt	SEQ ID NO 162	GGAGGCGATCACACCGCAGACCCTGATCAACATCCGGCCCGTCGTCGCCG
My480.txt	SEQ ID NO 163	GGAGGCGATCACACCGCAGACCCTGATCAACATCCGGCCCGTCGTCGCCG
MY212.txt	SEQ ID NO 164	GGAGGCGATCACACCGCAGACCCTGATCAACATCCGCCCGGTGGTCGCCG
MY491.txt	SEQ ID NO 165	GGAGGCGATCACACCGCAGACCTTGATCAACATCCGCCCGGTGGTCGCCG
MY497.txt	SEQ ID NO 166	GGAGGCGATCACACCGCAGACCCTGATCAACATCCGCCCGGTGGTCGCCG
MY816.txt	SEQ ID NO 167	GGAGGCGATCACACCGCAGACCCTGATCAACATCCGTCCCGTCGTGGCCG
Atcc10-xen.txt	SEQ ID NO 168	CCCAGGACGTGGAGGCGATCACACCGCAGACCTTGATCAACATCCGCCCCGTGGTGGCCG

Table 1

BASE NOS 1-60

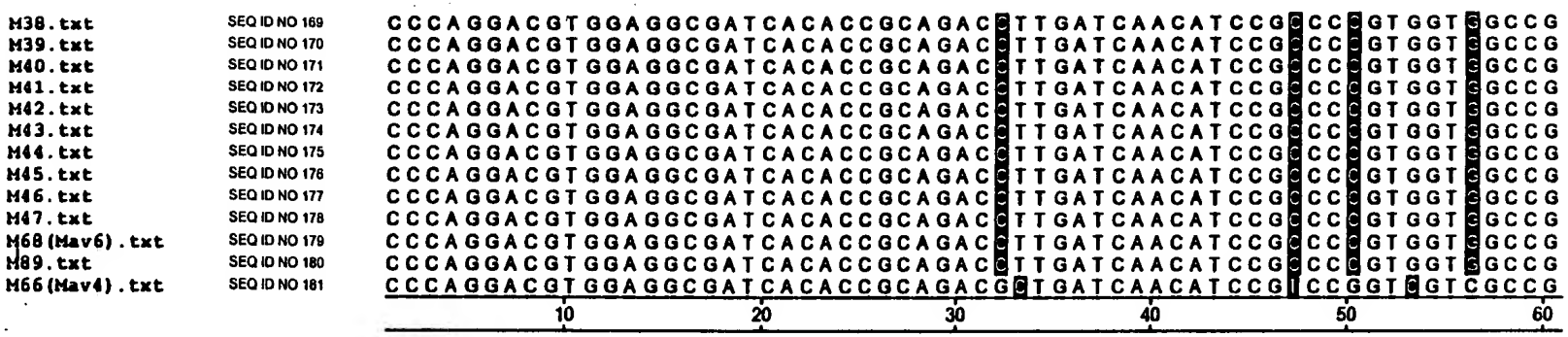

Table 1

BASE NOS 61-120

ATCC9-Mtb.txt	SEQ ID NO 1	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGAGCCAATTCATGGACCAGAACAACCCGC
MY621.txt	SEQ ID NO 2	CGATCAAGGAGTTCTTCGGAACCAGCCAGCTGTCGCAGTTCATGGACCAGAACAACCCGC
Atcc1-av.txt	SEQ ID NO 3	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCCAGTTCATGGACCAGAACAACCCGC
M29.txt	SEQ ID NO 4	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCCAGTTCATGGACCAGAACAACCCGC
M30.txt	SEQ ID NO 5	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCCAGTTCATGGACCAGAACAACCCGC
M31.txt	SEQ ID NO 6	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCCAGTTCATGGACCAGAACAACCCGC
M32.txt	SEQ ID NO 7	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCCAGTTCATGGACCAGAACAACCCGC
M33.txt	SEQ ID NO 8	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCCAGTTCATGGACCAGAACAACCCGC
M34.txt	SEQ ID NO 9	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCCAGTTCATGGACCAGAACAACCCGC
M48-new.txt	SEQ ID NO 10	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCCAGTTCATGGACCAGAACAACCCGC
M49.txt	SEQ ID NO 11	CGATCAAGGAGTTCTTCGGCACCAGCCAGTTGTCCAGTTCATGGACCAGAACAACCCGC
M64(Mav2).txt	SEQ ID NO 12	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCCAGTTCATGGACCAGAACAACCCGC
M65(Mav3).txt	SEQ ID NO 13	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCCAGTTCATGGACCAGAACAACCCGC
M67(Mav5).txt	SEQ ID NO 14	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCCAGTTCATGGACCAGAACAACCCGC
M69(Mav7).txt	SEQ ID NO 15	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCCAGTTCATGGACCAGAACAACCCGC
M71(Mav9).txt	SEQ ID NO 16	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCCAGTTCATGGACCAGAACAACCCGC
M91.txt	SEQ ID NO 17	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCCAGTTCATGGACCAGAACAACCCGC
M94.txt	SEQ ID NO 18	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCCAGTTCATGGACCAGAACAACCCGC
M95.txt	SEQ ID NO 19	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCCAGTTCATGGACCAGAACAACCCGC
M96.txt	SEQ ID NO 20	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCCAGTTCATGGACCAGAACAACCCGC
M100.txt	SEQ ID NO 21	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCCAGTTCATGGACCAGAACAACCCGC
M101.txt	SEQ ID NO 22	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCCAGTTCATGGACCAGAACAACCCGC
M102.txt	SEQ ID NO 23	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCCAGTTCATGGACCAGAACAACCCGC
M104.txt	SEQ ID NO 24	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCCAGTTCATGGACCAGAACAACCCGC
M105.txt	SEQ ID NO 25	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCCAGTTCATGGACCAGAACAACCCGC
M106.txt	SEQ ID NO 26	CGATCAAGGAATTCTTCGGCACCAGCCAGCTGTCGCAGTTCATGGACCAGAACAACCCGC
MY111.txt	SEQ ID NO 27	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCCCAGTTCATGGACCAGAACAACCCGC
M76.txt	SEQ ID NO 28	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGAGCCAATTCATGGACCAGAACAACCCGC
MY451.txt	SEQ ID NO 29	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGAGCCAATTCATGGACCAGAACAACCCGC
ATCC2-chelnew.txt	SEQ ID NO 30	CGATCAAGGAGTTCTTCGGAACCAGCCAGCTGTCGCAGTTCATGGACCAGAACAACCCGC
M10.txt	SEQ ID NO 31	CGATCAAGGAGTTCTTCGGAACCAGCCAGCTGTCGCAGTTCATGGACCAGAACAACCCGC
M11-662.txt	SEQ ID NO 32	CGATCAAGGAGTTCTTCGGAACCAGCCAGCTGTCGCAGTTCATGGACCAGAACAACCCGC
M12.txt	SEQ ID NO 33	CGATCAAGGAGTTCTTCGGAACCAGCCAGCTGTCGCAGTTCATGGACCAGAACAACCCGC
M13_2_662.txt	SEQ ID NO 34	CGATCAAGGAGTTCTTCGGAACCAGCCAGCTGTCGCAGTTCATGGACCAGAACAACCCGC
M14.txt	SEQ ID NO 35	CGATCAAGGAGTTCTTCGGAACCAGCCAGCTGTCGCAGTTCATGGACCAGAACAACCCGC
M15.txt	SEQ ID NO 36	CGATCAAGGAGTTCTTCGGAACCAGCCAGCTGTCGCAGTTCATGGACCAGAACAACCCGC
M16.txt	SEQ ID NO 37	CGATCAAGGAGTTCTTCGGAACCAGCCAGCTGTCGCAGTTCATGGACCAGAACAACCCGC
M17_2_662.txt	SEQ ID NO 38	CGATCAAGGAGTTCTTCGGAACCAGCCAGCTGTCGCAGTTCATGGACCAGAACAACCCGC
M50.txt	SEQ ID NO 39	CGATCAAGGAGTTCTTCGGAACCAGCCAGCTGTCGCAGTTCATGGACCAGAACAACCCGC
M51.txt	SEQ ID NO 40	CGATCAAGGAGTTCTTCGGAACCAGCCAGCTGTCGCAGTTCATGGACCAGAACAACCCGC
M115.txt	SEQ ID NO 41	CGATCAAGGAGTTCTTCGGAACCAGCCAGCTGTCGCAGTTCATGGACCAGAACAACCCGC
M116.txt	SEQ ID NO 42	CGATCAAGGAGTTCTTCGGAACCAGCCAGCTGTCGCAGTTCATGGACCAGAACAACCCGC
M119.txt	SEQ ID NO 43	CGATCAAGGAGTTCTTCGGAACCAGCCAGCTGTCGCAGTTCATGGACCAGAACAACCCGC
MY109.txt	SEQ ID NO 44	CGATCAAGGAGTTCTTCGGAACCAGCCAGCTGTCGCAGTTCATGGACCAGAACAACCCGC
MY200.txt	SEQ ID NO 45	CGATCAAGGAGTTCTTCGGAACCAGCCAGCTGTCGCAGTTCATGGACCAGAACAACCCGC
MY207.txt	SEQ ID NO 46	CGATCAAGGAGTTCTTCGGAACCAGCCAGCTGTCGCAGTTCATGGACCAGAACAACCCGC
MY209.txt	SEQ ID NO 47	CGATCAAGGAGTTCTTCGGAACCAGCCAGCTGTCGCAGTTCATGGACCAGAACAACCCGC
M122.txt	SEQ ID NO 48	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCGCAGTTCATGGACCAGAACAACCCGC
M123.txt	SEQ ID NO 49	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCGCAGTTCATGGACCAGAACAACCCGC
M124.txt	SEQ ID NO 50	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCGCAGTTCATGGAICAGAACAACCCGC
Atcc3-for.txt	SEQ ID NO 51	CGATCAAGGAGTTCTTCGGAACGTCGCAGCTGTCGCAGTTCATGGAICAGAACAACCCGC
M53.txt	SEQ ID NO 52	CGATCAAGGAATTCTTCGGCACCAGCCAGCTGTCGCAGTTCATGGACCAGAACAACCCGC
M54.txt	SEQ ID NO 53	CGATCAAGGAGTTCTTCGGTACCAGCCAGCTGTCGCAGTTCATGGACCAGAACAACCCGC
M55.txt	SEQ ID NO 54	CGATCAAGGAGTTCTTCGGAACGTCGCAGCTGTCGCAGTTCATGGAICAGAACAACCCGC
M56.txt	SEQ ID NO 55	CGATCAAGGAGTTCTTCGGCACCAGCCAGCTGTCGCAGTTCATGGACCAGAACAACCCGC
M74(Mav12).txt	SEQ ID NO 56	CGATCAAGGAGTTCTTCGGAACGTCGCAGCTGTCGCAGTTCATGGAICAGAACAACCCGC

Table 1

BASE NOS 61-120

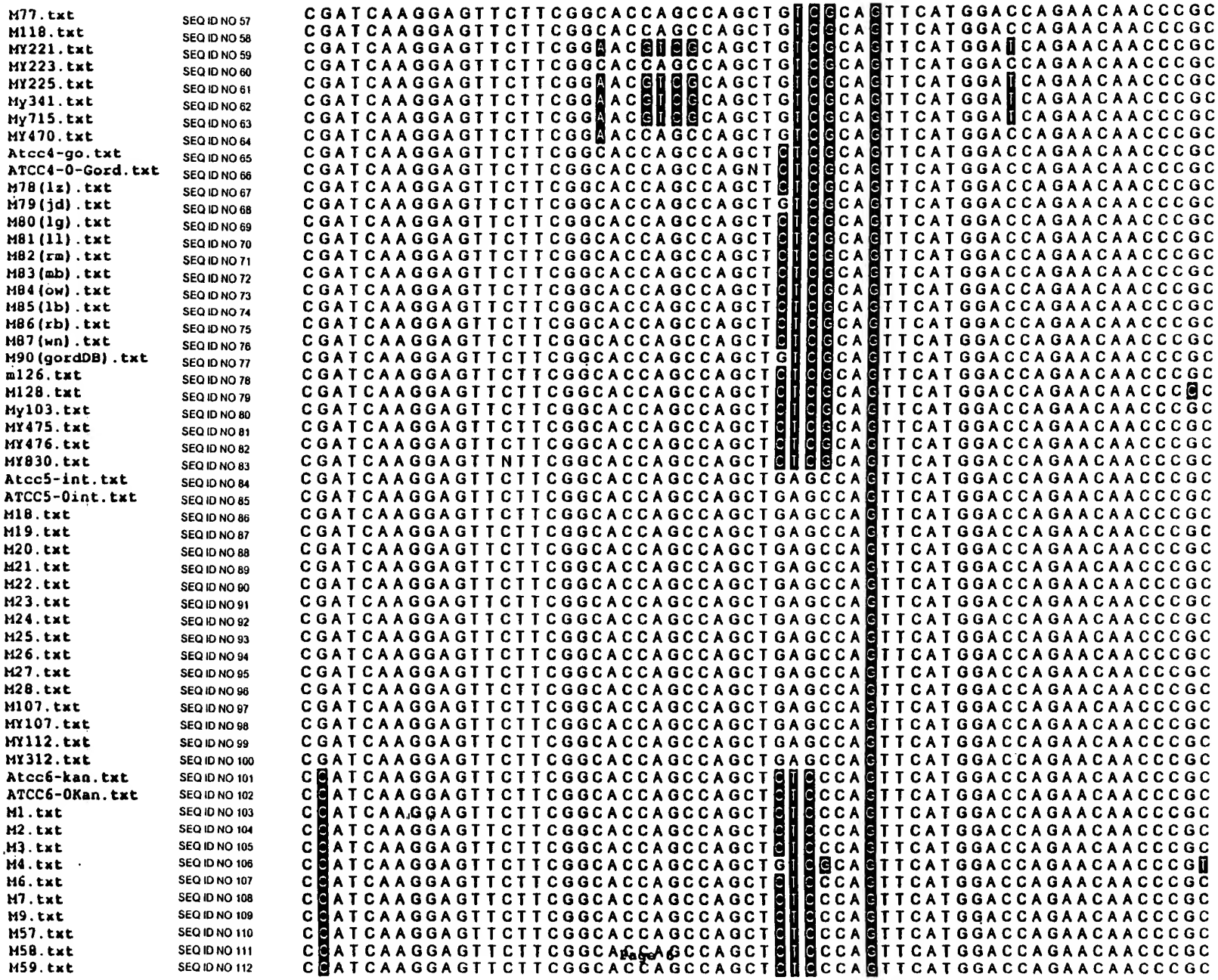

Table 1

BASE NOS 61-120

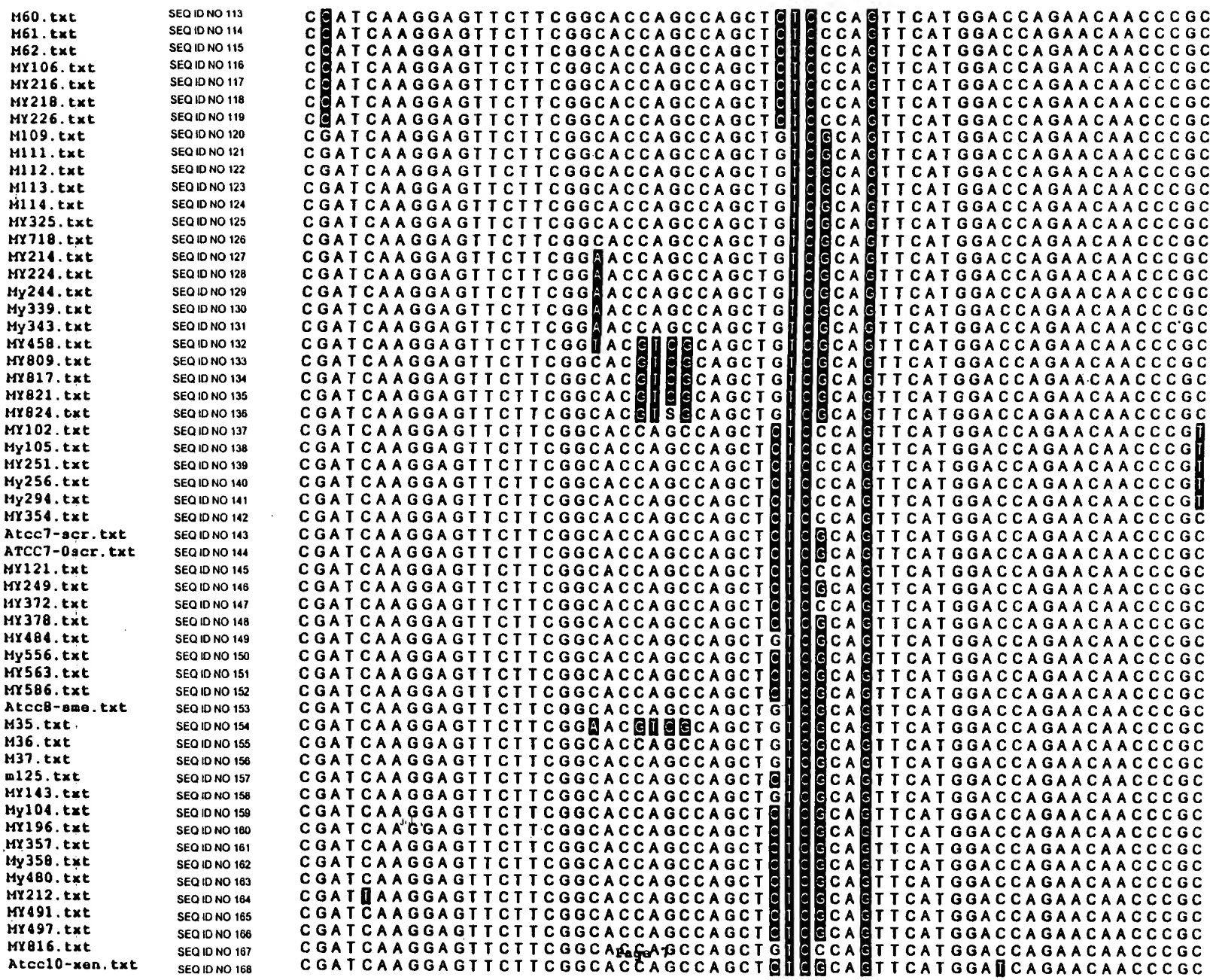

Table 1

BASE NOS 61-120

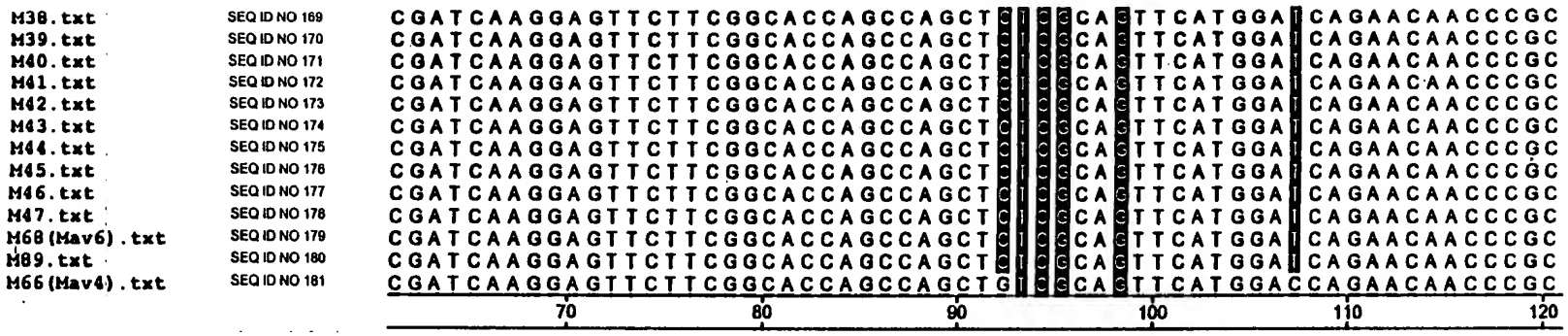

Table 1

BASE NOS 121-180

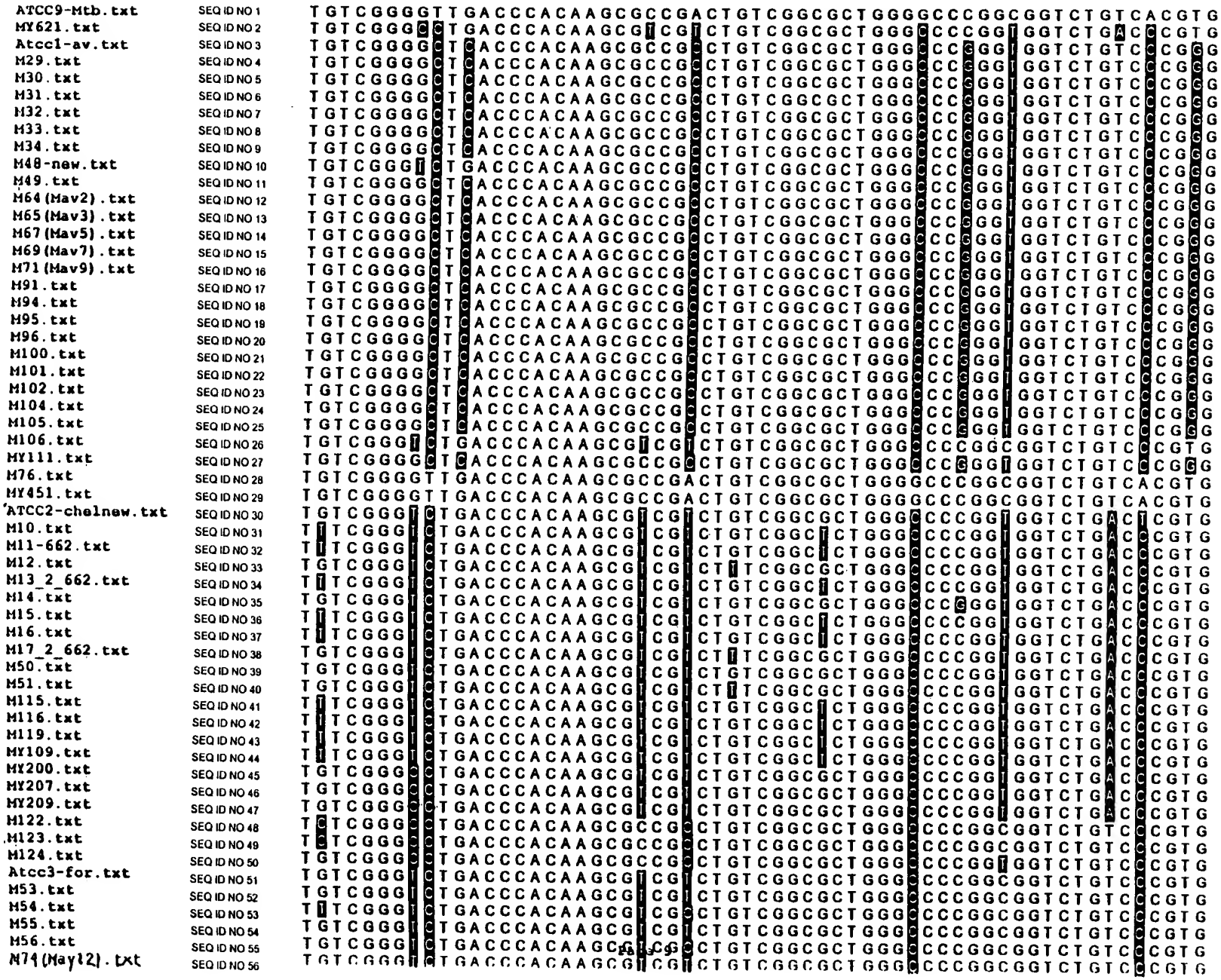

Table 1

BASE NOS 121-180

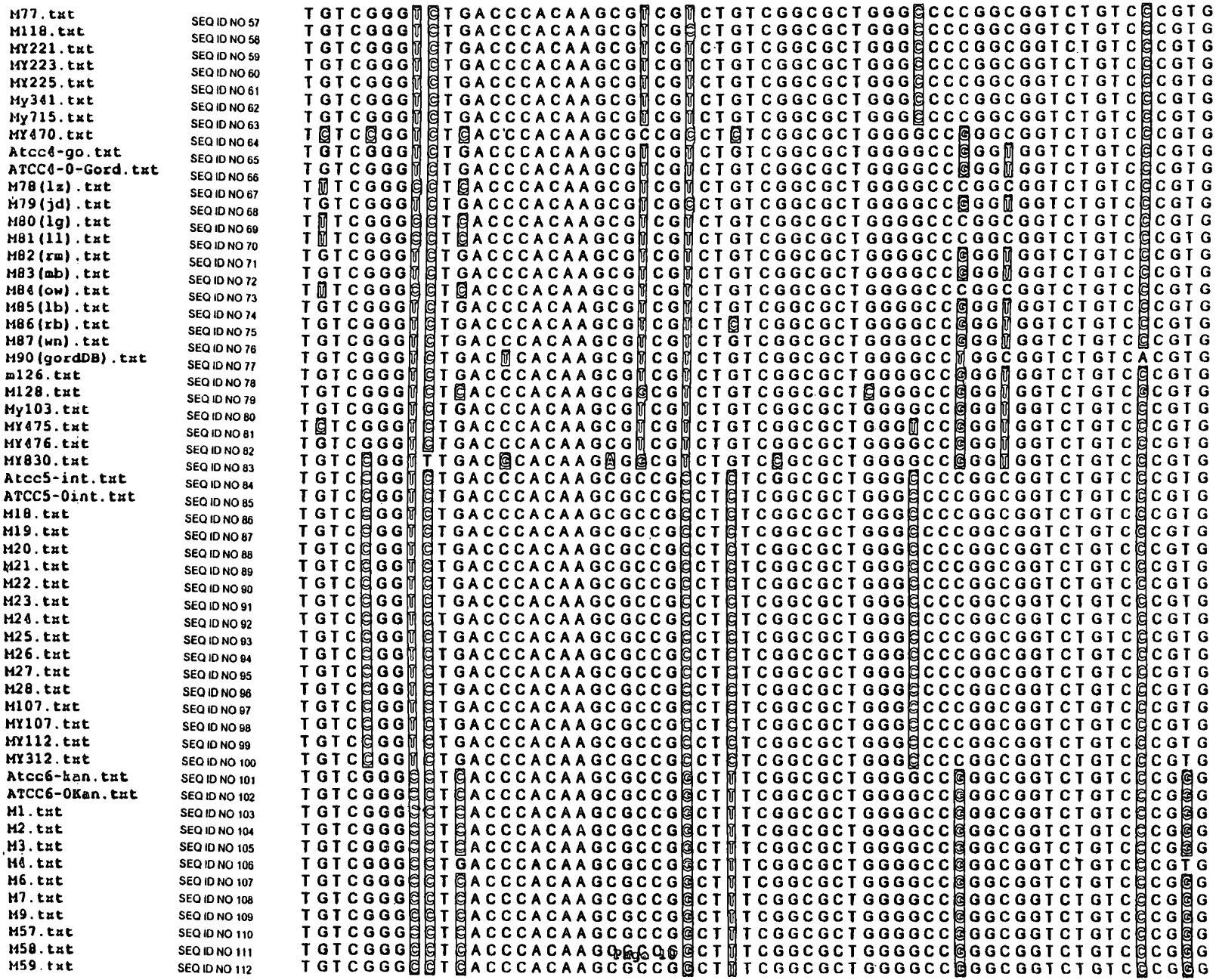

Table 1

BASE NOS 121-180

M60.txt	SEQ ID NO 113
M61.txt	SEQ ID NO 114
M62.txt	SEQ ID NO 115
MY106.txt	SEQ ID NO 116
MY216.txt	SEQ ID NO 117
MY218.txt	SEQ ID NO 118
MY226.txt	SEQ ID NO 119
M109.txt	SEQ ID NO 120
M111.txt	SEQ ID NO 121
M112.txt	SEQ ID NO 122
M113.txt	SEQ ID NO 123
M114.txt	SEQ ID NO 124
MY325.txt	SEQ ID NO 125
MY718.txt	SEQ ID NO 126
MY214.txt	SEQ ID NO 127
MY224.txt	SEQ ID NO 128
My244.txt	SEQ ID NO 129
My339.txt	SEQ ID NO 130
My343.txt	SEQ ID NO 131
MY458.txt	SEQ ID NO 132
MY809.txt	SEQ ID NO 133
MY817.txt	SEQ ID NO 134
MY821.txt	SEQ ID NO 135
MY824.txt	SEQ ID NO 136
MY102.txt	SEQ ID NO 137
My105.txt	SEQ ID NO 138
MY251.txt	SEQ ID NO 139
My256.txt	SEQ ID NO 140
My294.txt	SEQ ID NO 141
MY354.txt	SEQ ID NO 142
Atcc7-scr.txt	SEQ ID NO 143
ATCC7-0scr.txt	SEQ ID NO 144
MY121.txt	SEQ ID NO 145
MY249.txt	SEQ ID NO 146
MY372.txt	SEQ ID NO 147
MY378.txt	SEQ ID NO 148
MY484.txt	SEQ ID NO 149
My556.txt	SEQ ID NO 150
MY563.txt	SEQ ID NO 151
MY586.txt	SEQ ID NO 152
Atcc8-sme.txt	SEQ ID NO 153
M35.txt	SEQ ID NO 154
M36.txt	SEQ ID NO 155
M37.txt	SEQ ID NO 156
m125.txt	SEQ ID NO 157
MY143.txt	SEQ ID NO 158
My104.txt	SEQ ID NO 159
MY196.txt	SEQ ID NO 160
MY357.txt	SEQ ID NO 161
My358.txt	SEQ ID NO 162
My480.txt	SEQ ID NO 163
MY212.txt	SEQ ID NO 164
MY491.txt	SEQ ID NO 165
MY497.txt	SEQ ID NO 166
MY816.txt	SEQ ID NO 167
Atcc10-xen.txt	SEQ ID NO 168

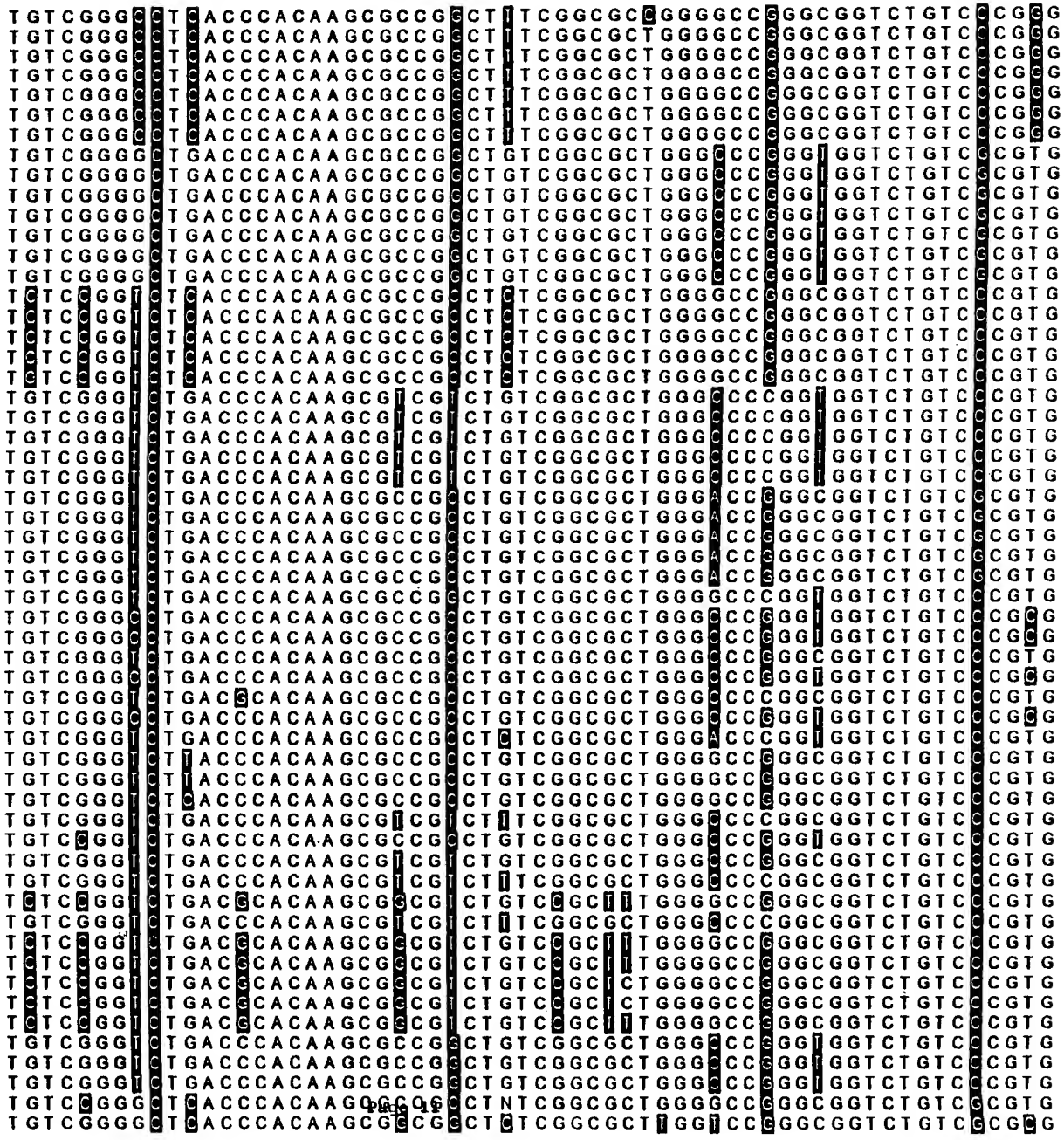

Table 1

BASE NOS 121-180

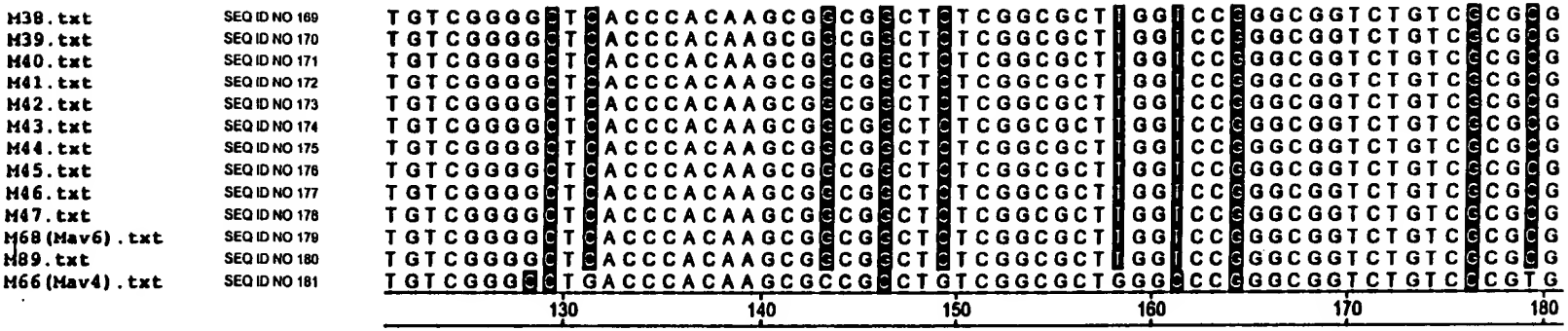

Table 1

BASE NOS 181-240

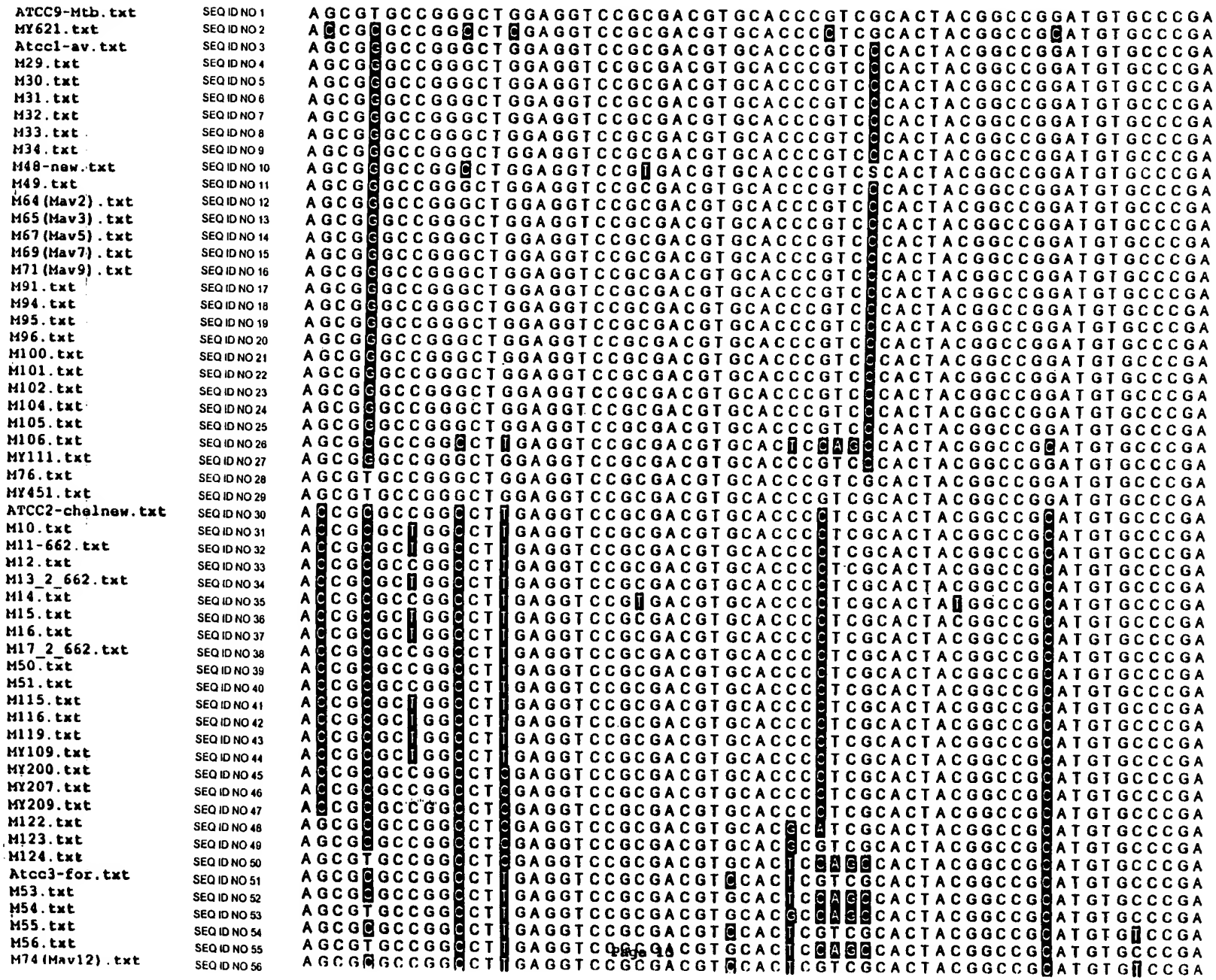

Table 1

BASE NOS 181-240

M77.txt	SEQ ID NO 57	AGCGCGCCGGCCTTGAGGTCCGCGACGTGCACTCCAGCCACTACGGCCGCATGTGCCCGA
M118.txt	SEQ ID NO 58	AGCGTGCCGGCCTTGAGGTCCGCGACGTGCACTCCAGCCACTACGGCCGCATGTGCCCGA
MY221.txt	SEQ ID NO 59	AGCGCGCCGGCCTTGAGGTCCGCGACGTGCACTCGTCGCACTACGGCCGCATGTGCCCGA
MY223.txt	SEQ ID NO 60	AGCGCGCCGGCCTTGAGGTCCGCGACGTGCACTCCAGCCACTACGGCCGCATGTGCCCGA
MY225.txt	SEQ ID NO 61	AGCGCGCCGGCCTTGAGGTCCGCGACGTCCACTCGTCGCACTACGGCCGCATGTGCCCGA
My341.txt	SEQ ID NO 62	AGCGCGCCGGCCTTGAGGTCCGCGACGTCCACTCGTCGCACTACGGCCGCATGTGCCCGA
My715.txt	SEQ ID NO 63	AGCGCGCCGGCCTTGAGGTCCGCGACGTCCACTCGTCGCACTACGGCCGCATGTGTCCGA
MY470.txt	SEQ ID NO 64	AGCGCGCCGGGTCTGGAAGTTCGTGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
Atcc4-go.txt	SEQ ID NO 65	AGCGTGCGGGTCTGGAAGTACGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA
ATCC4-0-Gord.txt	SEQ ID NO 66	AGCGTGCGGTGCTGGAAGTACGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA
M78(lz).txt	SEQ ID NO 67	AGCGGGCCGGCCTGGAGGTCCGTGACGTCCACCCGTCGCACTACGGCCGCATGTGCCCGA
M79(jd).txt	SEQ ID NO 68	AGCGTGCCGGCCTGGAGGTCCGTGACGTCCACCCGTCGCACTACGGCCGCATGTGCCCGA
M80(lg).txt	SEQ ID NO 69	AGCGGGCCGGCCTGGAGGTCCGTGACGTCCACCCGTCGCACTACGGCCGCATGTGCCCGA
M81(ll).txt	SEQ ID NO 70	AGCGGGCCGGCCTGGAGGTCCGTGACGTCCACCCGTCGCACTACGGCCGCATGTGCCCGA
M82(rm).txt	SEQ ID NO 71	AGCGTGCGGGTCTGGAAGTACGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA
M83(mb).txt	SEQ ID NO 72	AGCGTGCGGGTCTGGAAGTACGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA
M84(ow).txt	SEQ ID NO 73	AGCGGGCCGGCCTGGAGGTCCGTGACGTCCACCCGTCGCACTACGGCCGCATGTGCCCGA
M85(lb).txt	SEQ ID NO 74	AGCGTGCGGGTCTGGAAGTACGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA
M86(rb).txt	SEQ ID NO 75	AGCGCGCCGGGTCTGGAGGTCCGTGACGTCCACCCGTCGCACTACGGCCGCATGTGCCCGA
M87(wn).txt	SEQ ID NO 76	AGCGTGCGGGTCTGGAAGTACGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA
M90(gordDB).txt	SEQ ID NO 77	AGCGCGCCGGCCTGGAAGTCCGTGACGTCCACCCGTCGCACTACGGCCGGATGTGCCCGA
m126.txt	SEQ ID NO 78	AGCGTGCGGGTCTGGAAGTACGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA
M128.txt	SEQ ID NO 79	AGCGTGCGGGTCTGGAAGTCCGTGACGTCCACCCGTCGCACTACGGCCGCATGTGCCCGA
My103.txt	SEQ ID NO 80	AGCGTGCGGGTCTGGAAGTACGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA
MY475.txt	SEQ ID NO 81	AGCGCGCCGGGTCTGGAGGTCCGTGACGTCCACCCGTCGCACTACGGCCGCATGTGCCCGA
MY476.txt	SEQ ID NO 82	AGCGTGCGGGTCTGGAAGTACGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA
MY830.txt	SEQ ID NO 83	AGCGGGCCGGCCTGGAGGTCCGTGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
Atcc5-int.txt	SEQ ID NO 84	AGCGCGCCGGCCTGGAGGTCCGTGACGTCCACCCGTCGCACTACGGCCGGATGTGCCCGA
ATCC5-0int.txt	SEQ ID NO 85	AGCGCGCCGGCCTGGAGGTCCGTGACGTCCACCCGTCGCACTACGGCCGGATGTGCCCGA
M18.txt	SEQ ID NO 86	AGCGCGCCGGCCTGGAGGTCCGTGACGTCCACCCGTCGCACTACGGCCGGATGTGCCCGA
M19.txt	SEQ ID NO 87	AGCGCGCCGGCCTGGAGGTCCGTGACGTCCACCCGTCGCACTACGGCCGGATGTGCCCGA
M20.txt	SEQ ID NO 88	AGCGCGCCGGCCTGGAGGTCCGTGACGTCCACCCGTCGCACTACGGCCGGATGTGCCCGA
M21.txt	SEQ ID NO 89	AGCGCGCCGGCCTGGAGGTCCGTGACGTCCACCCGTCGCACTACGGCCGGATGTGCCCGA
M22.txt	SEQ ID NO 90	AGCGCGCCGGCCTGGAGGTCCGTGACGTCCACCCGTCGCACTACGGCCGGATGTGCCCGA
M23.txt	SEQ ID NO 91	AGCGCGCCGGCCTGGAGGTCCGTGACGTCCACCCGTCGCACTACGGCCGGATGTGCCCGA
M24.txt	SEQ ID NO 92	AGCGCGCCGGCCTGGAGGTCCGTGACGTCCACCCGTCGCACTACGGCCGGATGTGCCCGA
M25.txt	SEQ ID NO 93	AGCGCGCCGGCCTGGAGGTCCGTGACGTCCACCCGTCGCACTACGGCCGGATGTGCCCGA
M26.txt	SEQ ID NO 94	AGCGCGCCGGCCTGGAGGTCCGTGACGTCCACCCGTCGCACTACGGCCGGATGTGCCCGA
M27.txt	SEQ ID NO 95	AGCGCGCCGGCCTGGAGGTCCGTGACGTCCACCCGTCGCACTACGGCCGGATGTGCCCGA
M28.txt	SEQ ID NO 96	AGCGCGCCGGCCTGGAGGTCCGTGACGTCCACCCGTCGCACTACGGCCGGATGTGCCCGA
M107.txt	SEQ ID NO 97	AGCGCGCCGGCCTGGAGGTCCGTGACGTCCACCCGTCGCACTACGGCCGGATGTGCCCGA
MY107.txt	SEQ ID NO 98	AGCGCGCCGGCCTGGAGGTCCGTGACGTCCACCCGTCGCACTACGGCCGGATGTGCCCGA
MY112.txt	SEQ ID NO 99	AGCGCGCCGGCCTGGAGGTCCGTGACGTCCACCCGTCGCACTACGGCCGGATGTGCCCGA
MY312.txt	SEQ ID NO 100	AGCGCGCCGGCCTGGAGGTCCGTGACGTCCACCCGTCGCACTACGGCCGGATGTGCCCGA
Atcc6-kan.txt	SEQ ID NO 101	AGCGTGCCGGGCTGGAAGTGCGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA
ATCC6-0Kan.txt	SEQ ID NO 102	AGCGTGCCGGGCTGGAAGTTCGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA
M1.txt	SEQ ID NO 103	AGCGTGCCGGGCTGGAAGTGCGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA
M2.txt	SEQ ID NO 104	AGCGTGCCGGGCTGGAAGTGCGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA
M3.txt	SEQ ID NO 105	AGCGTGCCGGGCTGGAAGTGCGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA
M4.txt	SEQ ID NO 106	AGCGTGCCGGCCTGGAAGTGCGTGACGTGCACCCTTCGCACTACGGCCGGATGTGCCCGA
M6.txt	SEQ ID NO 107	AGCGTGCCGGGCTGGAAGTGCGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA
M7.txt	SEQ ID NO 108	AGCGTGCCGGGCTGGAAGTGCGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA
M9.txt	SEQ ID NO 109	AGCGTGCCGGGCTGGAAGTGCGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA
M57.txt	SEQ ID NO 110	AGCGTGCCGGGCTGGAAGTGCGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA
M58.txt	SEQ ID NO 111	AGCGTGCCGGGCTGGAAGTGCGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA
M59.txt	SEQ ID NO 112	AGCGTGCCGGGCTGGAAGTGCGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA

Table 1

BASE NOS 181-240

M60.txt	SEQ ID NO 113	AGCGTGCCGGGCTGGAAGTGCGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA
M61.txt	SEQ ID NO 114	AGCGTGCCGGGCTGGAAGTGCGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA
M62.txt	SEQ ID NO 115	AGCGTGCCGGGCTGGAAGTGCGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA
MY106.txt	SEQ ID NO 116	AGCGTGCCGGGCTGGAAGTGCGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA
MY216.txt	SEQ ID NO 117	AGCGTGCCGGGCTGGAAGTGCGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA
MY218.txt	SEQ ID NO 118	AGCGTGCCGGGCTGGAAGTGCGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA
MY226.txt	SEQ ID NO 119	AGCGTGCCGGGCTGGAAGTGCGTGACGTGCACCCGTCGCACTACGGCCGCATGTGCCCGA
M109.txt	SEQ ID NO 120	AGCGTGCCGGCTTGGAGGTCCGTGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
M111.txt	SEQ ID NO 121	AGCGTGCCGGCTTGGAGGTCCGTGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
M112.txt	SEQ ID NO 122	AGCGTGCCGGCTTGGAGGTCCGTGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
M113.txt	SEQ ID NO 123	AGCGTGCCGGCTTGGAGGTCCGTGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
M114.txt	SEQ ID NO 124	AGCGTGCCGGCTTGGAGGTCCGTGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
MY325.txt	SEQ ID NO 125	AGCGTGCCGGCTTGGAGGTCCGTGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
MY718.txt	SEQ ID NO 126	AGCGTGCCGGCTTGGAGGTCCGTGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
MY214.txt	SEQ ID NO 127	AGCGCGCCGGTCTGGAAGTTCGTGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
MY224.txt	SEQ ID NO 128	AGCGCGCCGGTCTGGAAGTTCGTGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
My244.txt	SEQ ID NO 129	AGCGCGCCGGTCTGGAAGTTCGTGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
My339.txt	SEQ ID NO 130	AGCGCGCCGGTCTGGAAGTTCGTGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
My343.txt	SEQ ID NO 131	AGCGCGCCGGTCTGGAAGTTCGTGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
MY458.txt	SEQ ID NO 132	AGCGCGCCGGCCTCGAGGTCCGCGACGTCCACTCGTCGCACTACGGCCGCATGTGCCCGA
MY809.txt	SEQ ID NO 133	AGCGCGCCGGCCTCGAGGTCCGCGACGTCCACTCGTCGCACTACGGCCGCATGTGCCCGA
MY817.txt	SEQ ID NO 134	AGCGCGCCGGCCTCGAGGTCCGCGACGTCCACTCGTCGCACTACGGCCGCATGTGCCCGA
MY821.txt	SEQ ID NO 135	AGCGCGCCGGCCTCGAGGTCCGCGACGTCCACTCGTCGCACTACGGCCGCATGTGCCCGA
MY824.txt	SEQ ID NO 136	AGCGCGCCGGCCTCGAGGTYCGCGACGTCCACTCGTCGCACTACGGCCGCATGTGCCCGA
MY102.txt	SEQ ID NO 137	AGCGGGCCGGCCTGGAAGTTCGTGACGTGCACCCGTCCCACTACGGCCGGATGTGTCCGA
My105.txt	SEQ ID NO 138	AGCGGGCCGGCCTGGAAGTTCGTGACGTGCACCCGTCCCACTACGGCCGGATGTGTCCGA
MY251.txt	SEQ ID NO 139	AGCGGGCCGGCCTGGAAGTTCGTGACGTGCACCCGTCCCACTACGGCCGGATGTGTCCGA
My256.txt	SEQ ID NO 140	AGCGGGCCGGCCTGGAAGTTCGTGACGTGCACCCGTCCCACTACGGCCGGATGTGTCCGA
My294.txt	SEQ ID NO 141	AGCGGGCCGGCCTGGAAGTTCGTGACGTGCACCCGAGCCACTACGGCCGGATGTGTCCGA
MY354.txt	SEQ ID NO 142	AGCGGGCCGGCCTGGAAGTTCGTGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
Atcc7-scr.txt	SEQ ID NO 143	AGCGGGCCGGGCTGGAGGTCCGGGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
ATCC7-0scr.txt	SEQ ID NO 144	AGCGGGCCGGGCTGGAGGTCCGGGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
MY121.txt	SEQ ID NO 145	AGCGGGCCGGCCTCGAGGTGCGCGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
MY249.txt	SEQ ID NO 146	AGCGGGCCGGGCTGGAGGTCCGGGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
MY372.txt	SEQ ID NO 147	AGCGGGCCGGGCTGGAGGTCCGCGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
MY378.txt	SEQ ID NO 148	AGCGGGCCGGGCTGGAGGTCCGGGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
MY484.txt	SEQ ID NO 149	AGCGTGCGGGCCTCGAGGTCCGTGACGTACACCCGTCGCACTACGGCCGGATGTGTCCGA
My556.txt	SEQ ID NO 150	AGCGGGCGGGCCTCGAGGTCCGCGATGTGCACCCGTCGCACTACGGCCGGATGTGTCCGA
MY563.txt	SEQ ID NO 151	AGCGGGCGGGCCTCGAGGTCCGCGATGTGCACCCGTCGCACTACGGCCGGATGTGTCCGA
MY586.txt	SEQ ID NO 152	AGCGGGCGGGCCTCGAGGTCCGAGACGTGCACCCGTCGCACTACGGCCGGATGTGTCCGA
Atcc8-sme.txt	SEQ ID NO 153	AGCGCGCTGGCCTCGAGGTCCGCGACGTGCACCCCAGCCACTACGGCCGCATGTGCCCGA
M35.txt	SEQ ID NO 154	AGCGCGCCGGCCTGGAGGTCCGCGACGTGCACTCCAGCCACTACGGCCGGATGTGCCCGA
M36.txt	SEQ ID NO 155	AGCGCGCCGGCCTGGAGGTCCGCGACGTGCACTCCAGCCACTACGGCCGGATGTGCCCGA
M37.txt	SEQ ID NO 156	AGCGCGCCGGCCTCGAGGTCCGCGACGTGCACCCCAGCCACTACGGCCGCATGTGCCCGA
m125.txt	SEQ ID NO 157	AGCGGGCCGGGCTGGAGGTCCGTGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
MY143.txt	SEQ ID NO 158	AGCGGGCCGGGCCTCGAGGTCCGCGACGTGCACCCCAGCCACTACGGCCGCATGTGCCCGA
My104.txt	SEQ ID NO 159	AGCGGGCCCGGGCTGGAGGTCCGTGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
MY196.txt	SEQ ID NO 160	AGCGGGCCGGGCTGGAGGTCCGTGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
MY357.txt	SEQ ID NO 161	AGCGGGCCGGGCTGGAGGTCCGTGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
My358.txt	SEQ ID NO 162	AGCGGGCCGGGCTGGAGGTCCGTGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
My480.txt	SEQ ID NO 163	AGCGGGCCGGGCTGGAGGTCCGTGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
MY212.txt	SEQ ID NO 164	AACGGGCCGGGCTTGAGGTCCGTGACGTGCACCCGTCCCACTACGGCCGGATGTGTCCGA
MY491.txt	SEQ ID NO 165	AGCGGGCCGGGCTTGAGGTCCGTGACGTGCACCCGTCCCACTACGGCCGGATGTGTCCGA
MY497.txt	SEQ ID NO 166	AGCGGGCCGGGCTTGAGGTCCGTGACGTGCACCCGTCCCACTACGGCCGGATGTGTCCGA
MY816.txt	SEQ ID NO 167	AGCGCGCCGGGCCTCGAGGTTCGCGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGA
Atcc10-xen.txt	SEQ ID NO 168	AGCGGGCCGGGCTGGAGGTCCGTGACGTGCACTCGAGCCACTACGGCCGGATGTGCCCGA

Table 1

BASE NOS 181-240

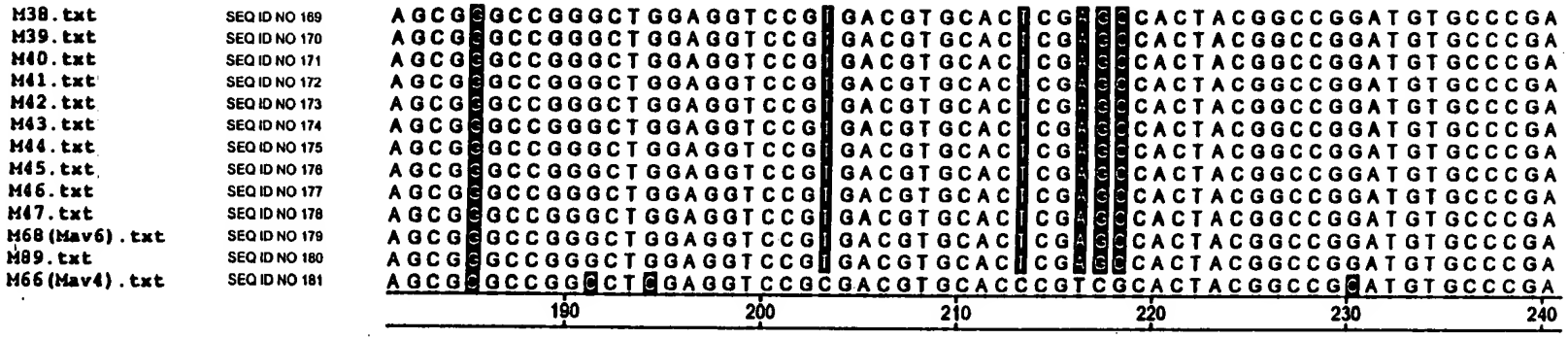

Table 1

BASE NOS 241-300

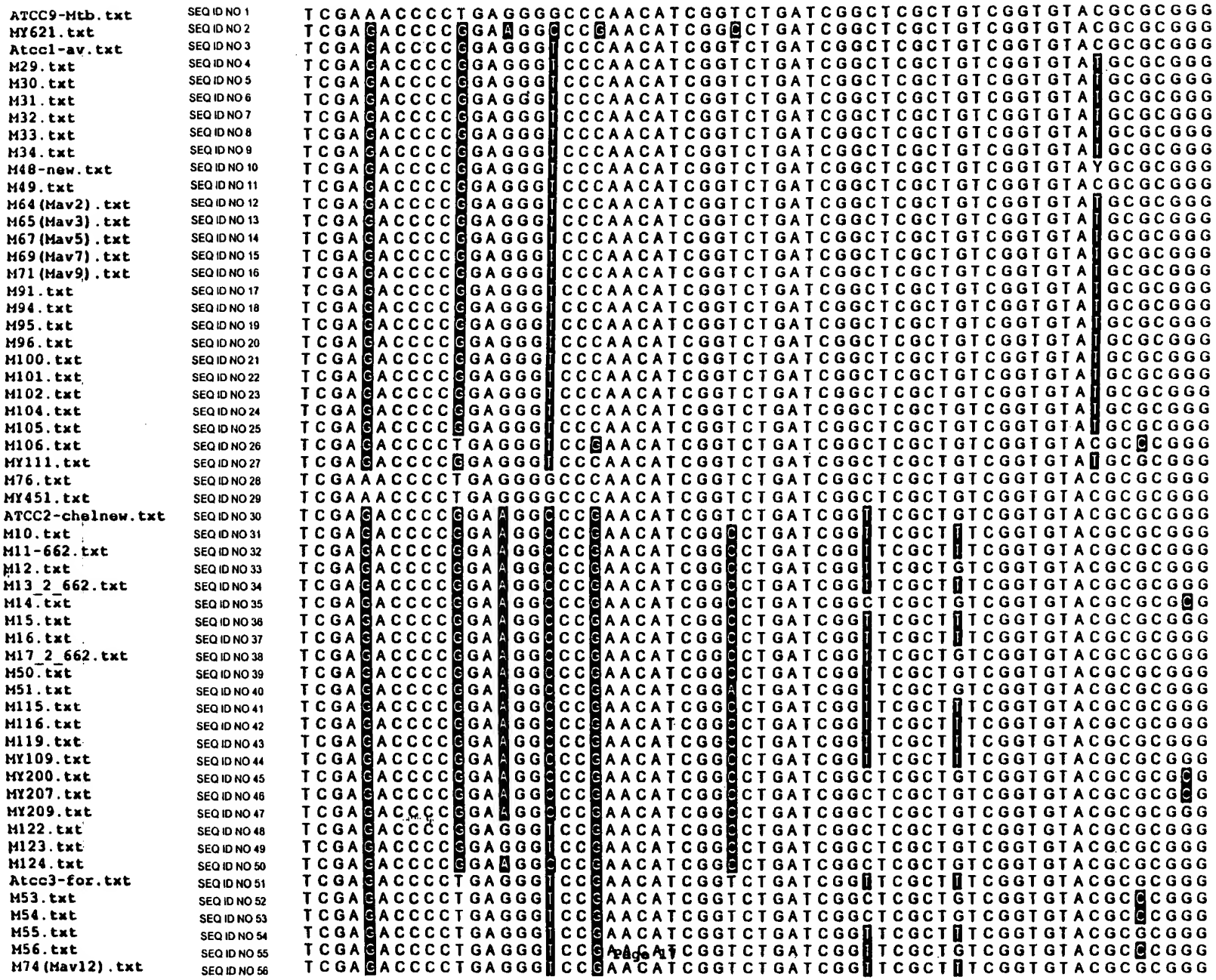

Table 1

BASE NOS 241-300

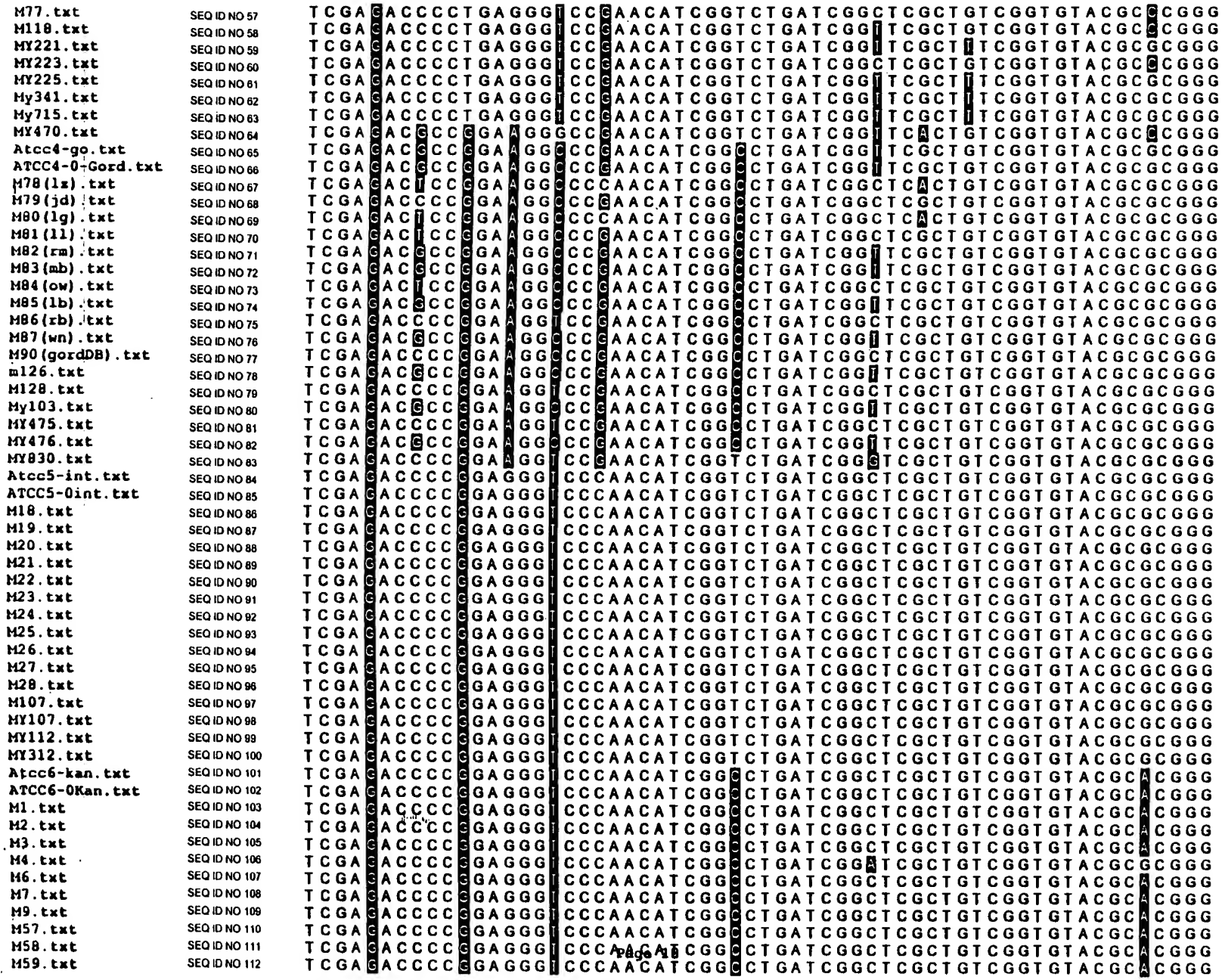

Table 1

BASE NOS 241-300

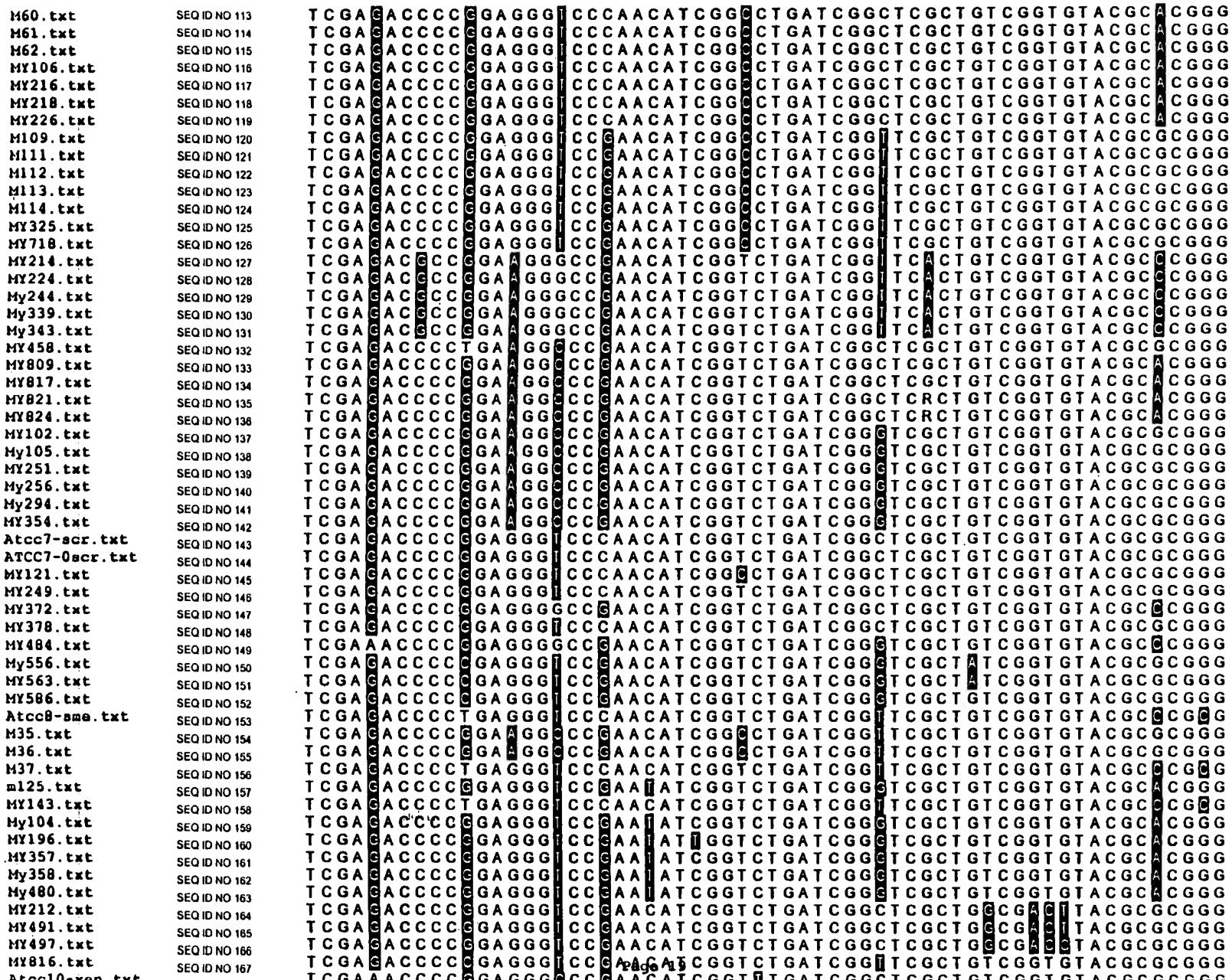

Table 1

BASE NOS 241-300

M38.txt	SEQ ID NO 169	TCGAAACCCCGGAGGGGCCGAACATCGGTTTGATCGGCTCGCTGTCGGTGTACGCGCGGG
M39.txt	SEQ ID NO 170	TCGAAACCCCGGAGGGGCCGAACATCGGTTTGATCGGCTCGCTGTCGGTGTACGCGCGGG
M40.txt	SEQ ID NO 171	TCGAAACCCCGGAGGGGCCGAACATCGGTTTGATCGGCTCGCTGTCGGTGTACGCGCGGG
M41.txt	SEQ ID NO 172	TCGAAACCCCGGAGGGGCCGAACATCGGTTTGATCGGCTCGCTGTCGGTGTACGCGCGGG
M42.txt	SEQ ID NO 173	TCGAAACCCCGGAGGGGCCGAACATCGGTTTGATCGGCTCGCTGTCGGTGTACGCGCGGG
M43.txt	SEQ ID NO 174	TCGAAACCCCGGAGGGGCCGAACATCGGTTTGATCGGCTCGCTGTCGGTGTACGCGCGGG
M44.txt	SEQ ID NO 175	TCGAAACCCCGGAGGGGCCGAACATCGGTTTGATCGGCTCGCTGTCGGTGTACGCGCGGG
M45.txt	SEQ ID NO 176	TCGAAACCCCGGAGGGGCCGAACATCGGTTTGATCGGCTCGCTGTCGGTGTACGCGCGGG
M46.txt	SEQ ID NO 177	TCGAAACCCCGGAGGGGCCGAACATCGGTTTGATCGGCTCGCTGTCGGTGTACGCGCGGG
M47.txt	SEQ ID NO 178	TCGAAACCCCGGAGGGGCCGAACATCGGTTTGATCGGCTCGCTGTCGGTGTACGCGCGGG
M68(Mav6).txt	SEQ ID NO 179	TCGAAACCCCGGAGGGGCCGAACATCGGTTTGATCGGCTCGCTGTCGGTGTACGCGCGGG
M89.txt	SEQ ID NO 180	TCGAAACCCCGGAGGGGCCGAACATCGGTTTGATCGGCTCGCTGTCGGTGTACGCGCGGG
M66(Mav4).txt	SEQ ID NO 181	TCGAGACCCCGGAGGGTCCGAACATCGGTCTGATCGGCTCGCTGTCGGTGTACGCGAGGG

250 260 270 280 290 300

Table 1

BASE NOS 301-360

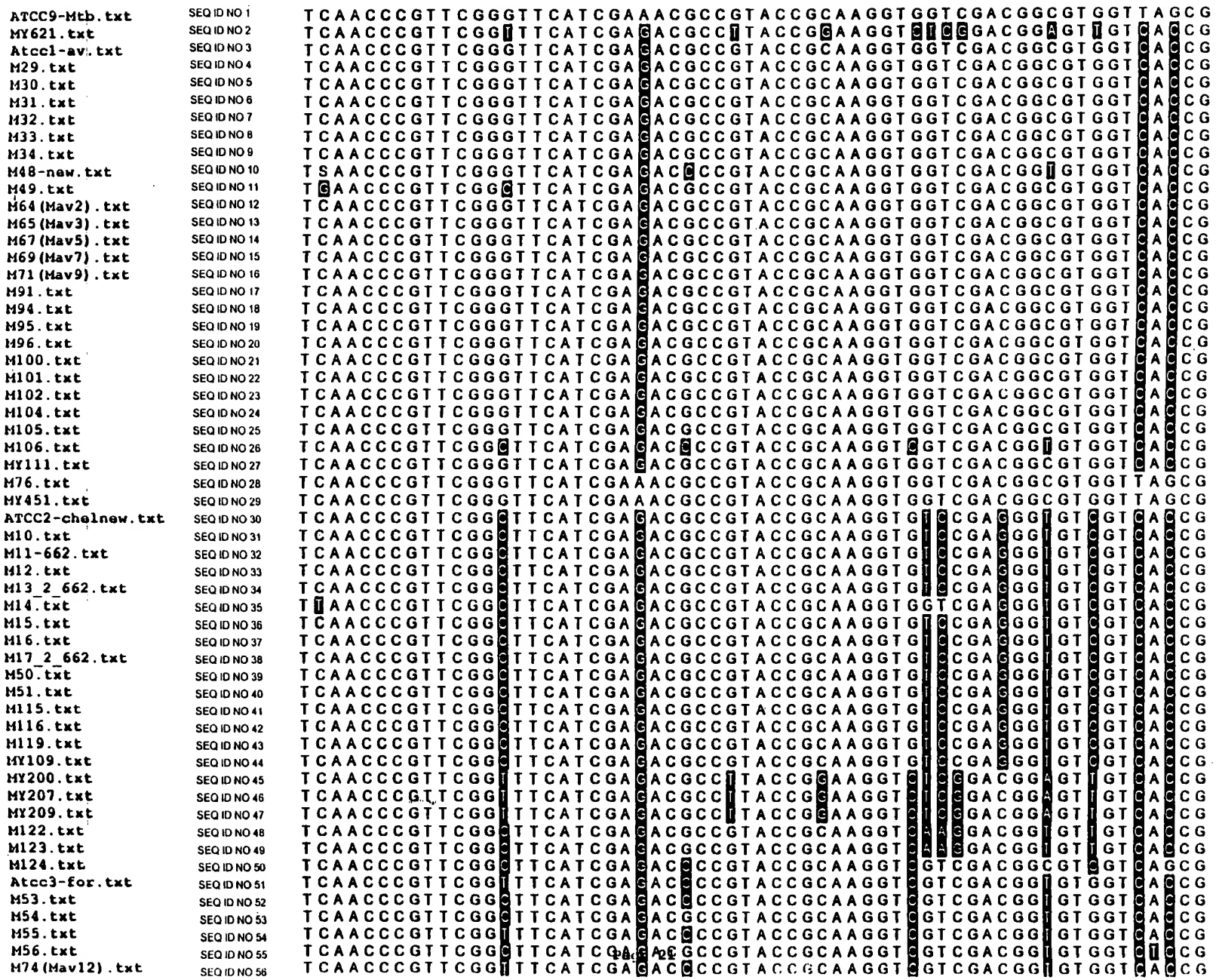

Table 1

BASE NOS 301-360

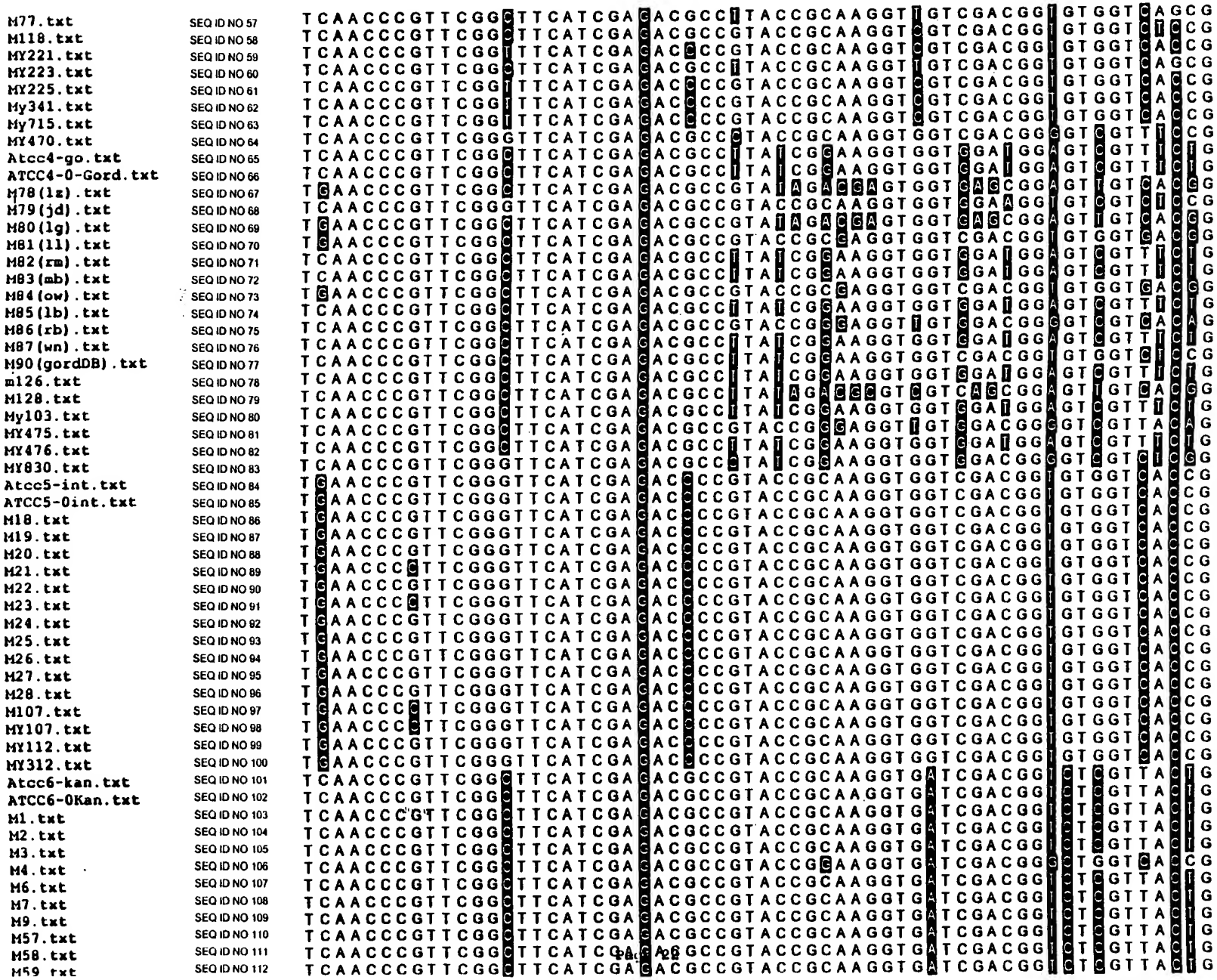

Table 1

BASE NOS 301-360

Table 1

BASE NOS 301-360

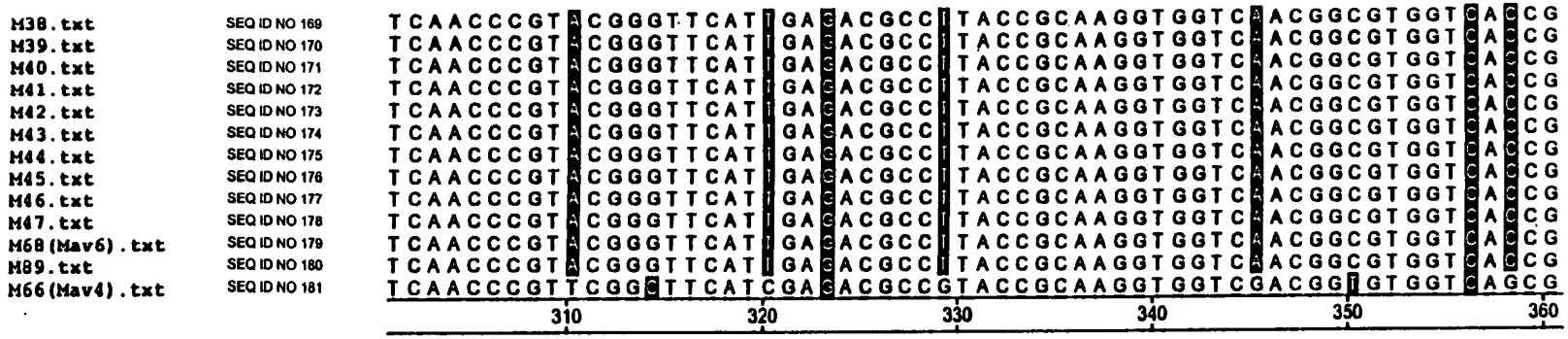

Table 1

BASE NOS 361-420

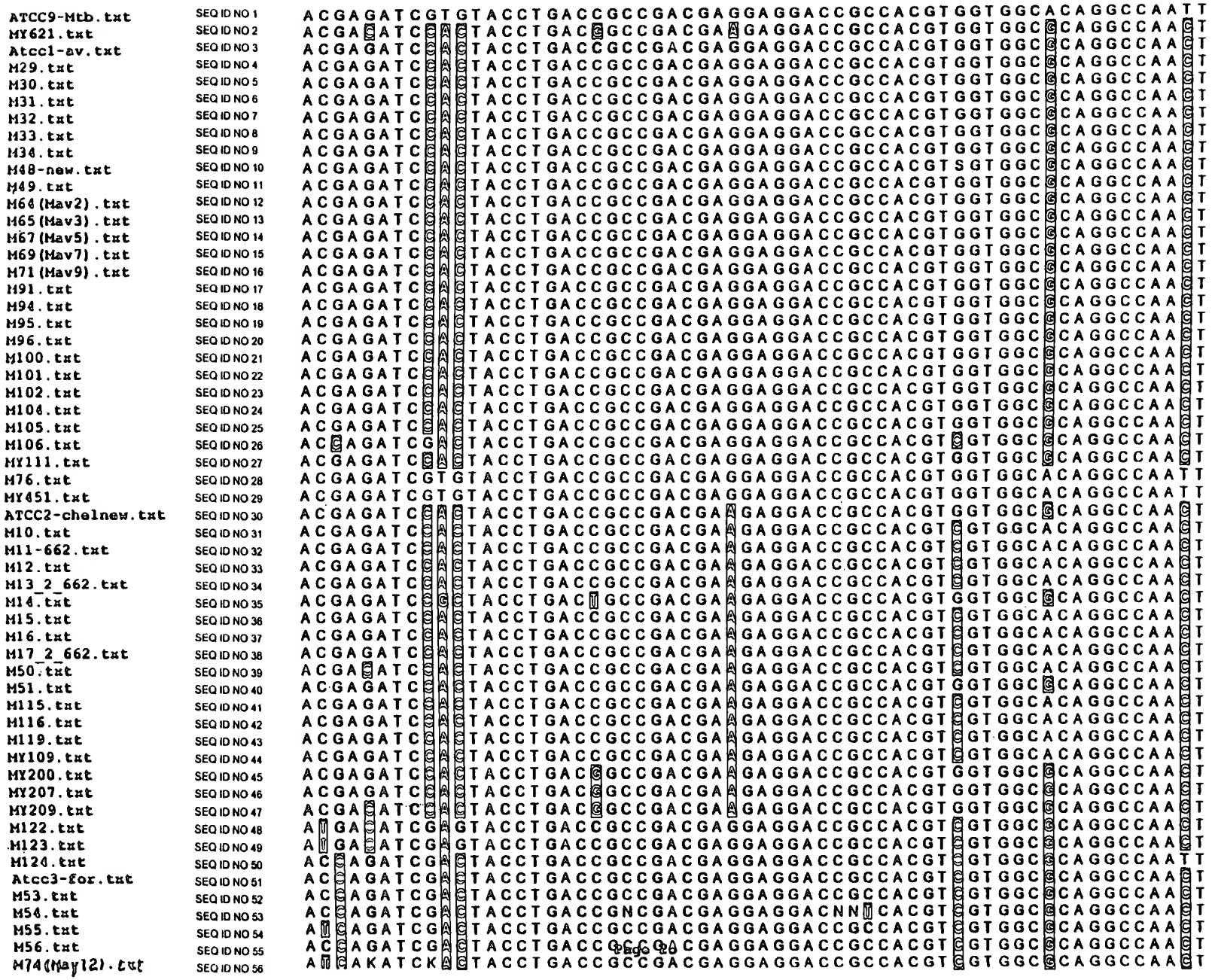

Table 1

BASE NOS 361-420

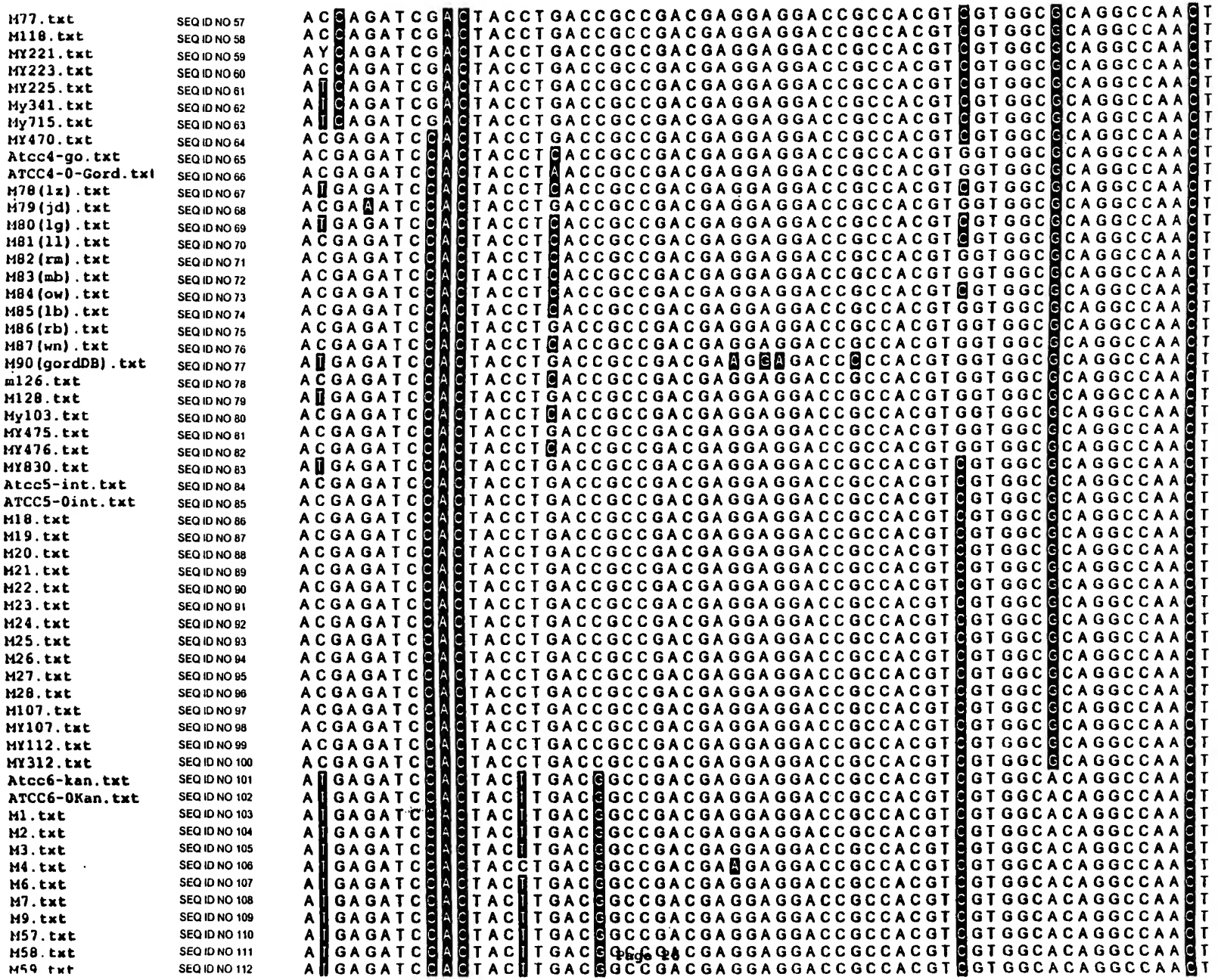

Table 1

BASE NOS 361-420

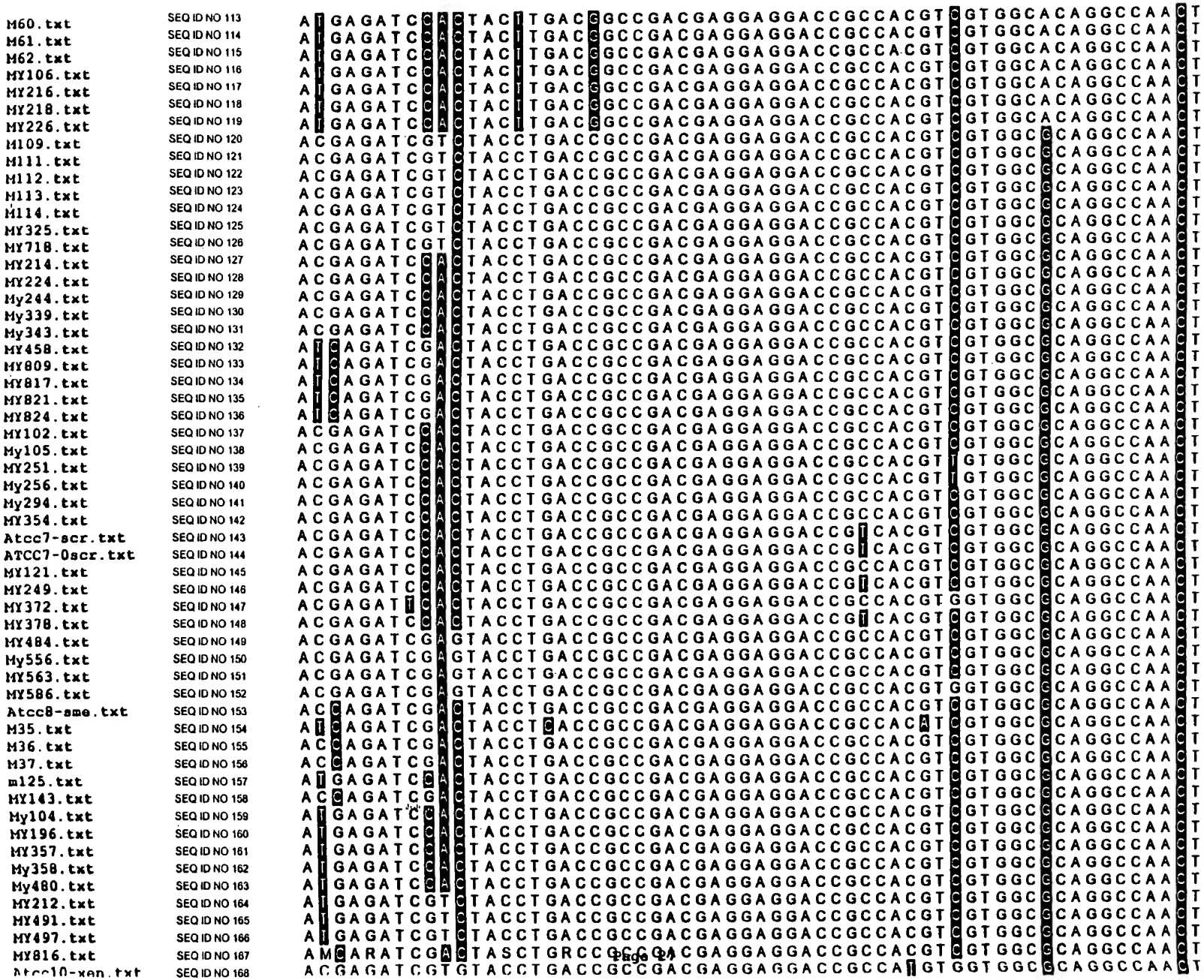

Table 1

BASE NOS 361-420

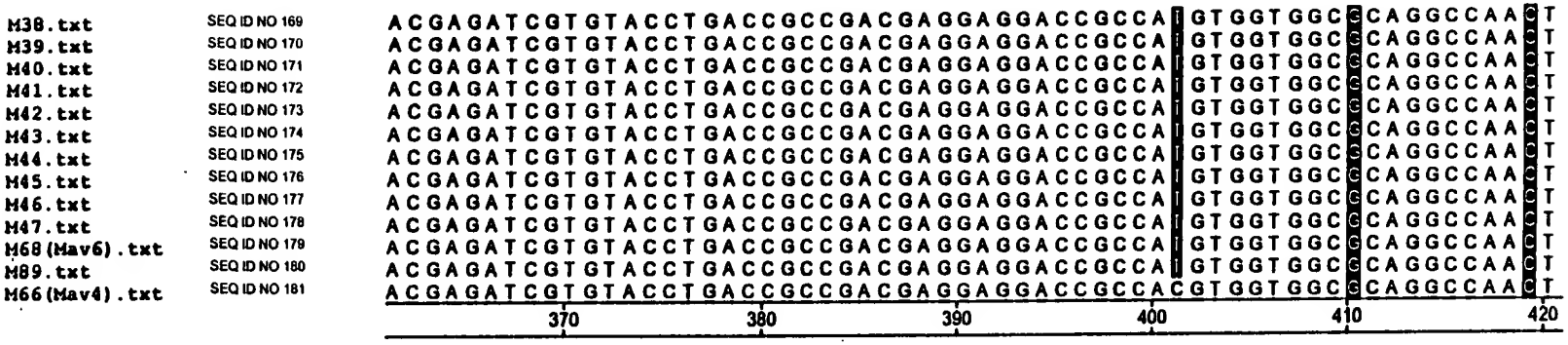

Table 1

BASE NOS 421-480

ATCC9-Mtb.txt	SEQ ID NO 1	CGCCGATC
MY621.txt	SEQ ID NO 2	CGCCCGTG
Atcc1-av.txt	SEQ ID NO 3	CGCCGATC
M29.txt	SEQ ID NO 4	CGCCGATC
M30.txt	SEQ ID NO 5	CGCCGATC
M31.txt	SEQ ID NO 6	CGCCGATC
M32.txt	SEQ ID NO 7	CGCCGATC
M33.txt	SEQ ID NO 8	CGCCGATC
M34.txt	SEQ ID NO 9	CGCCGATC
M48-new.txt	SEQ ID NO 10	CGCCGATC
M49.txt	SEQ ID NO 11	CGCCGATC
M64(Mav2).txt	SEQ ID NO 12	CGCCGATC
M65(Mav3).txt	SEQ ID NO 13	CGCCGATC
M67(Mav5).txt	SEQ ID NO 14	CGCCGATC
M69(Mav7).txt	SEQ ID NO 15	CGCCGATC
M71(Mav9).txt	SEQ ID NO 16	CGCCGATC
M91.txt	SEQ ID NO 17	CGCCGATC
M94.txt	SEQ ID NO 18	CGCCGATC
M95.txt	SEQ ID NO 19	CGCCGATC
M96.txt	SEQ ID NO 20	CGCCGATC
M100.txt	SEQ ID NO 21	CGCCGATC
M101.txt	SEQ ID NO 22	CGCCGATC
M102.txt	SEQ ID NO 23	CGCCGATC
M104.txt	SEQ ID NO 24	CGCCGATC
M105.txt	SEQ ID NO 25	CGCCGATC
M106.txt	SEQ ID NO 26	CGCCGATC
MY111.txt	SEQ ID NO 27	CGCCGATC
M76.txt	SEQ ID NO 28	CGCCGATC
MY451.txt	SEQ ID NO 29	CGCCGATC
ATCC2-chelnew.txt	SEQ ID NO 30	CGCCTGTG
M10.txt	SEQ ID NO 31	CGCCTGTG
M11-662.txt	SEQ ID NO 32	CGCCTGTG
M12.txt	SEQ ID NO 33	CGCCTGTG
M13_2_662.txt	SEQ ID NO 34	CGCCTGTG
M14.txt	SEQ ID NO 35	CGCCGACC
M15.txt	SEQ ID NO 36	CGCCTGTG
M16.txt	SEQ ID NO 37	CGCCTGTG
M17_2_662.txt	SEQ ID NO 38	CGCCTGTG
M50.txt	SEQ ID NO 39	CGCCTGTG
M51.txt	SEQ ID NO 40	CGCCTGTG
M115.txt	SEQ ID NO 41	CGCCTGTG
M116.txt	SEQ ID NO 42	CGCCTGTG
M119.txt	SEQ ID NO 43	CGCCTGTG
MY109.txt	SEQ ID NO 44	CGCCTGTG
MY200.txt	SEQ ID NO 45	CGCCCGTG
MY207.txt	SEQ ID NO 46	CGCCCGTG
MY209.txt	SEQ ID NO 47	CGCCCGTG
M122.txt	SEQ ID NO 48	CGCCGATC
M123.txt	SEQ ID NO 49	CGCCGATC
M124.txt	SEQ ID NO 50	CGCCGCTC
Atcc3-for.txt	SEQ ID NO 51	CGCCGATC
M53.txt	SEQ ID NO 52	CGCCGATC
M54.txt	SEQ ID NO 53	CGCCGATC
M55.txt	SEQ ID NO 54	CGCCGATC
M56.txt	SEQ ID NO 55	CGCCGATC
M74(May12).txt	SEQ ID NO 56	CGCCGATC

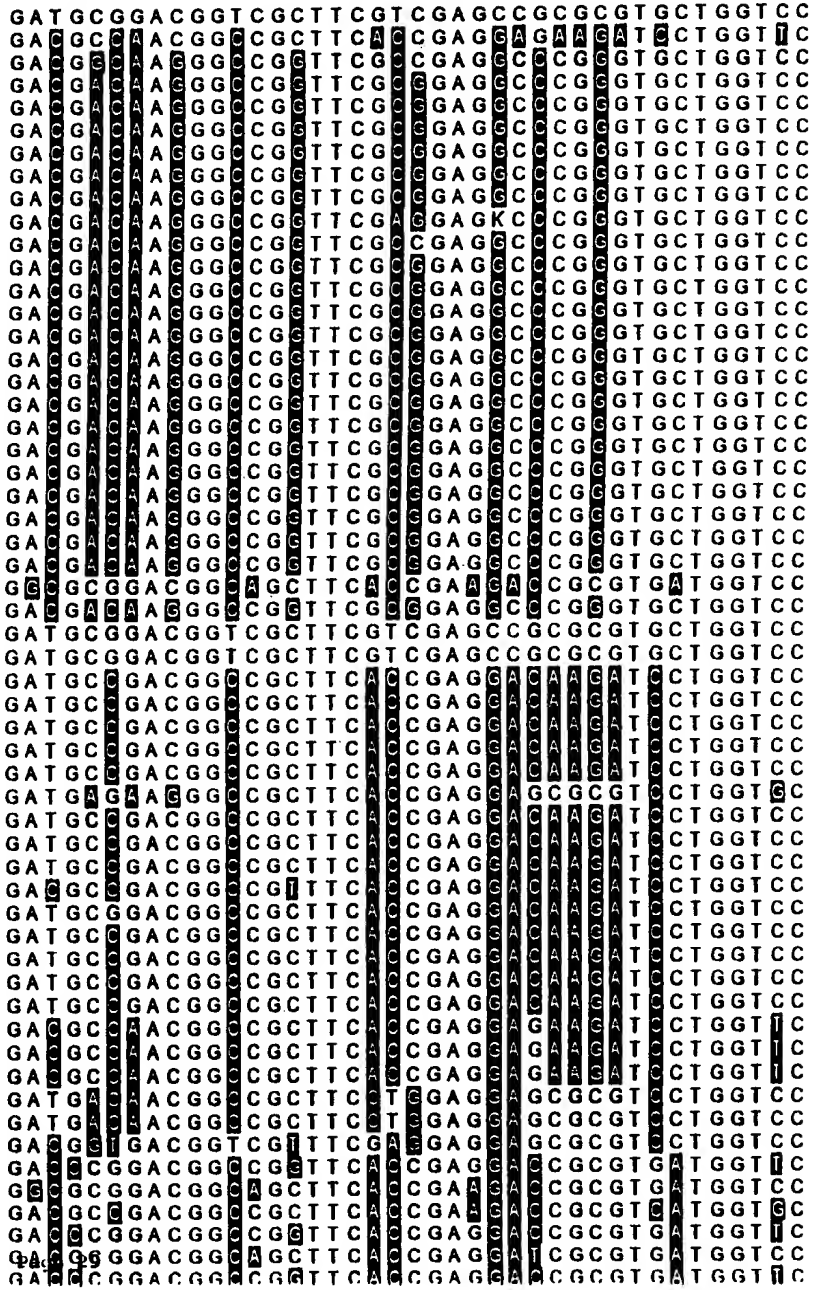

Table 1

BASE NOS 421-480

M77.txt		CGCCGATC
M118.txt	SEQ ID NO 57	CGCCGATC
MY221.txt	SEQ ID NO 58	CGCCGATC
MY223.txt	SEQ ID NO 59	CGCCGATC
MY225.txt	SEQ ID NO 60	CGCCGATC
My341.txt	SEQ ID NO 61	CGCCGATC
My715.txt	SEQ ID NO 62	CGCCGATC
MY470.txt	SEQ ID NO 63	CGCCGATC
Atcc4-go.txt	SEQ ID NO 64	CGCCGATC
ATCC4-0-Gord.txt	SEQ ID NO 65	CGCCGATC
M78(lz).txt	SEQ ID NO 66	CGCCGATC
M79(jd).txt	SEQ ID NO 67	CGCCGATC
M80(lg).txt	SEQ ID NO 68	CGCCGATC
M81(ll).txt	SEQ ID NO 69	CGCCGATC
M82(rm).txt	SEQ ID NO 70	CGCCGATC
M83(mb).txt	SEQ ID NO 71	CGCCGATC
M84(ow).txt	SEQ ID NO 72	CGCCGATC
M85(lb).txt	SEQ ID NO 73	CGCCGATC
M86(rb).txt	SEQ ID NO 74	CGCCGATC
M87(wn).txt	SEQ ID NO 75	CGCCGATC
M90(gordDB).txt	SEQ ID NO 76	CGCCGATC
m126.txt	SEQ ID NO 77	CGCCGATC
M128.txt	SEQ ID NO 78	CGCCGCTG
My103.txt	SEQ ID NO 79	CGCCGATC
MY475.txt	SEQ ID NO 80	CGCCGATC
MY476.txt	SEQ ID NO 81	CGCCGATC
MY830.txt	SEQ ID NO 82	CGCCGATC
Atcc5-int.txt	SEQ ID NO 83	CGCCGATC
ATCC5-0int.txt	SEQ ID NO 84	CGCCGATC
M18.txt	SEQ ID NO 85	CGCCGATC
M19.txt	SEQ ID NO 86	CGCCGATC
M20.txt	SEQ ID NO 87	CGCCGATC
M21.txt	SEQ ID NO 88	CGCCGATC
M22.txt	SEQ ID NO 89	CGCCGATC
M23.txt	SEQ ID NO 90	CGCCGATC
M24.txt	SEQ ID NO 91	CGCCGATC
M25.txt	SEQ ID NO 92	CGCCGATC
M26.txt	SEQ ID NO 93	CGCCGATC
M27.txt	SEQ ID NO 94	CGCCGATC
M28.txt	SEQ ID NO 95	CGCCGATC
M107.txt	SEQ ID NO 96	CGCCGATC
MY107.txt	SEQ ID NO 97	CGCCGATC
MY112.txt	SEQ ID NO 98	CGCCGATC
MY312.txt	SEQ ID NO 99	CGCCGATC
Atcc6-kan.txt	SEQ ID NO 100	CGCCGATC
ATCC6-0Kan.txt	SEQ ID NO 101	CGCCGATC
M1.txt	SEQ ID NO 102	CGCCGATC
M2.txt	SEQ ID NO 103	CGCCGATC
M3.txt	SEQ ID NO 104	CGCCGATC
M4.txt	SEQ ID NO 105	CGCCGATC
M6.txt	SEQ ID NO 106	CGCCGATC
M7.txt	SEQ ID NO 107	CGCCGATC
M9.txt	SEQ ID NO 108	CGCCGATC
M57.txt	SEQ ID NO 109	CGCCGATC
M58.txt	SEQ ID NO 110	CGCCGATC
M59.txt	SEQ ID NO 111	CGCCGATC
	SEQ ID NO 112	CGCCGATC

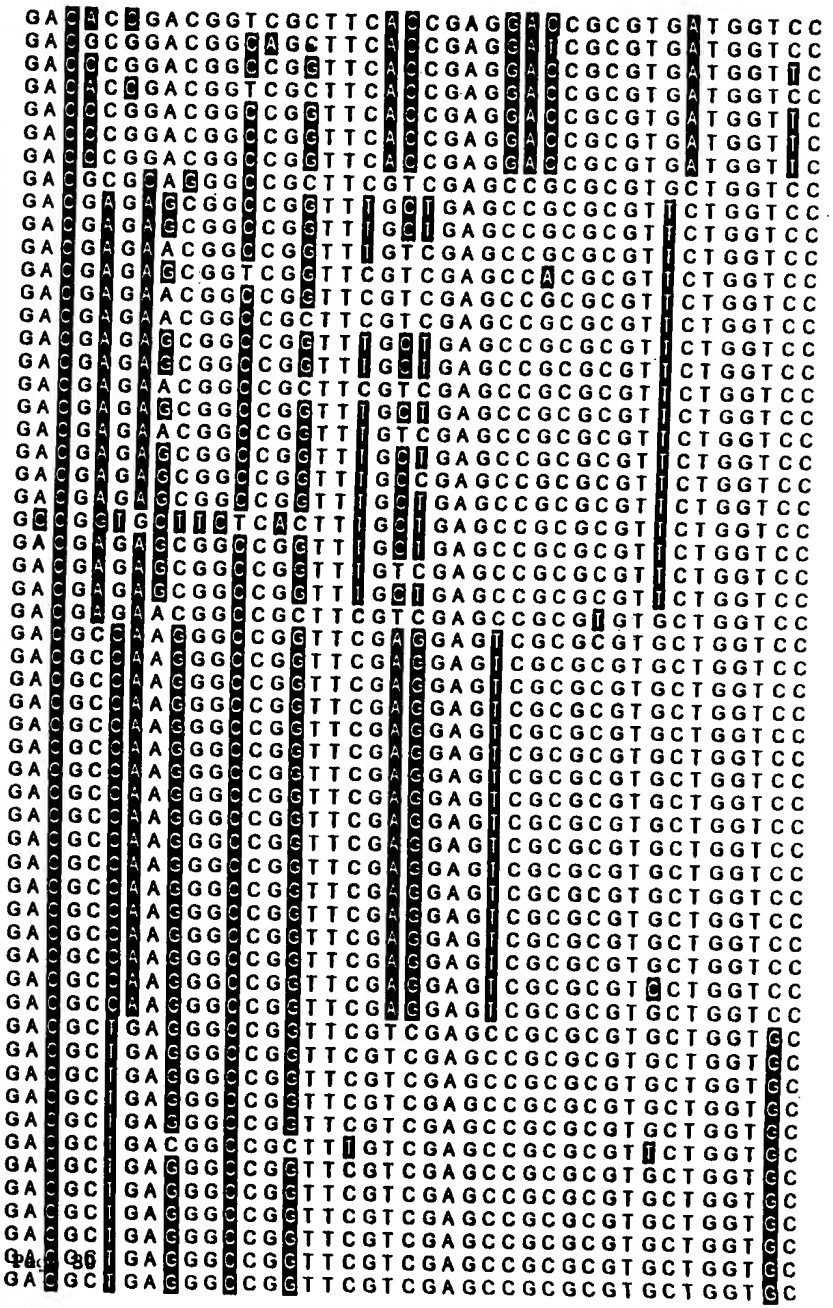

Table 1

BASE NOS 421-480

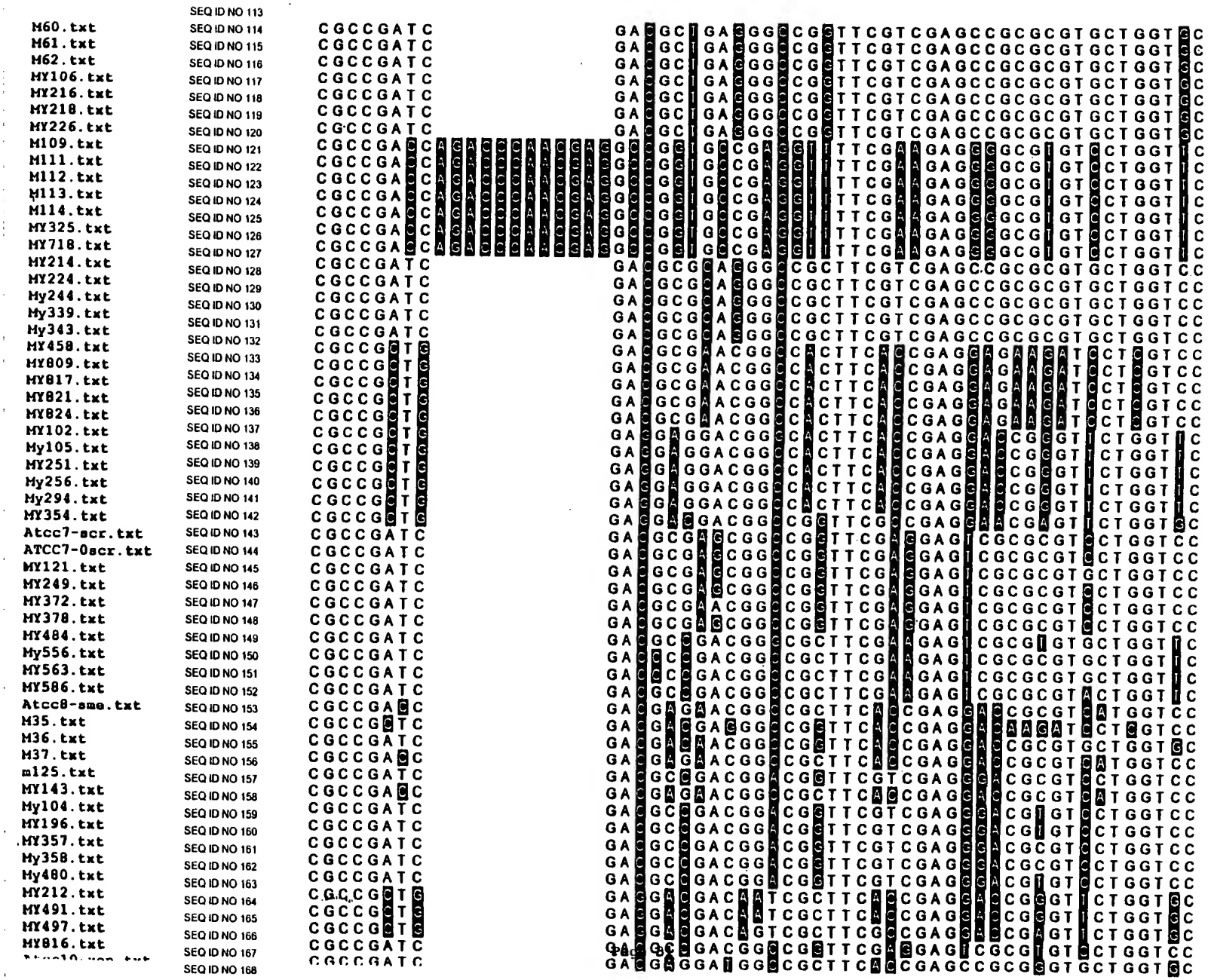

Table 1

BASE NOS 421-480

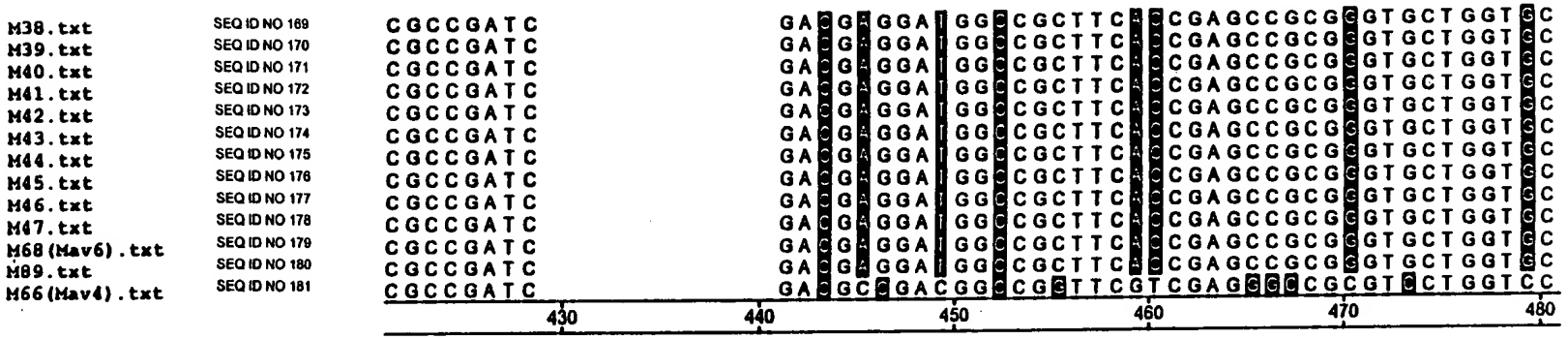

Table 1

BASE NOS 481-540

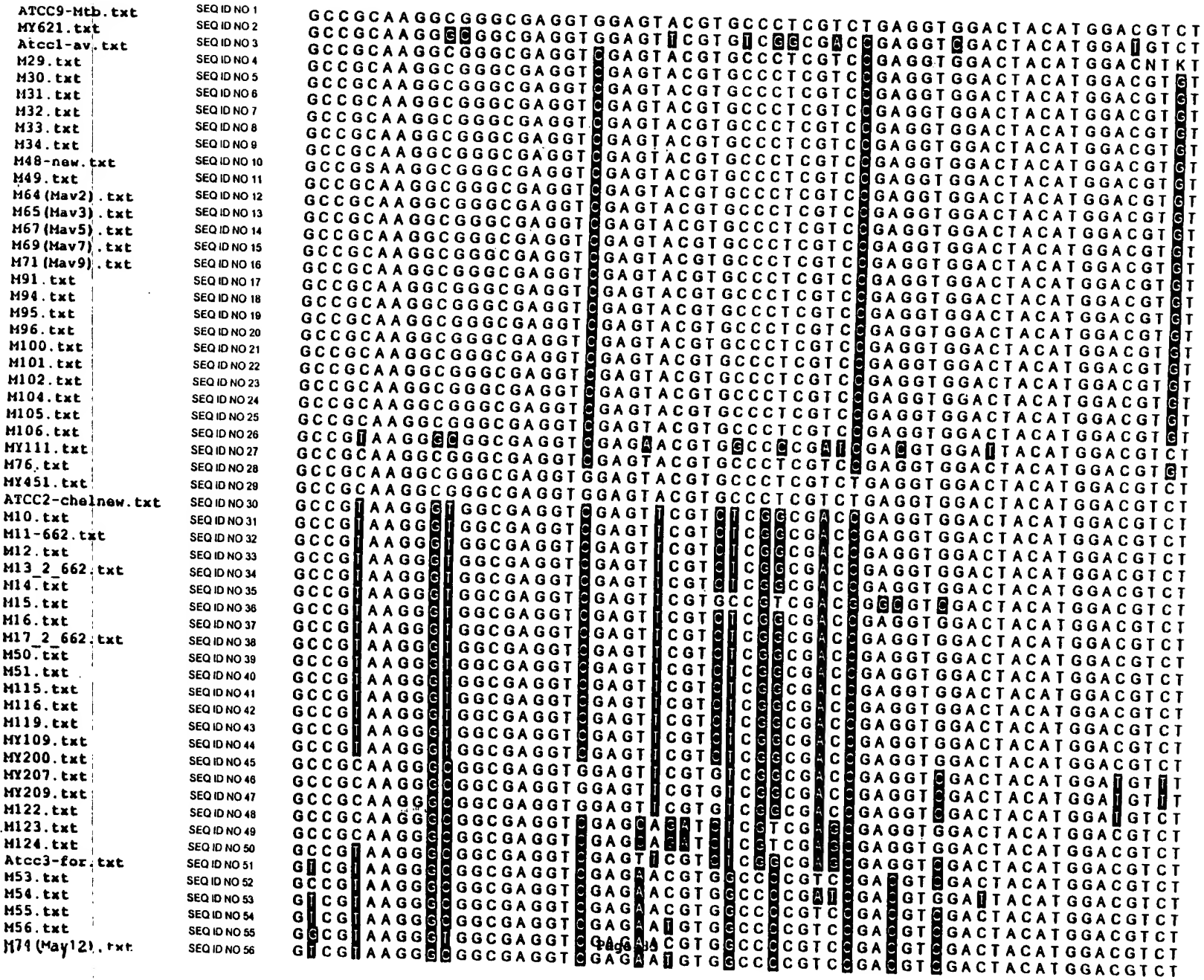

Table 1

BASE NOS 481-540

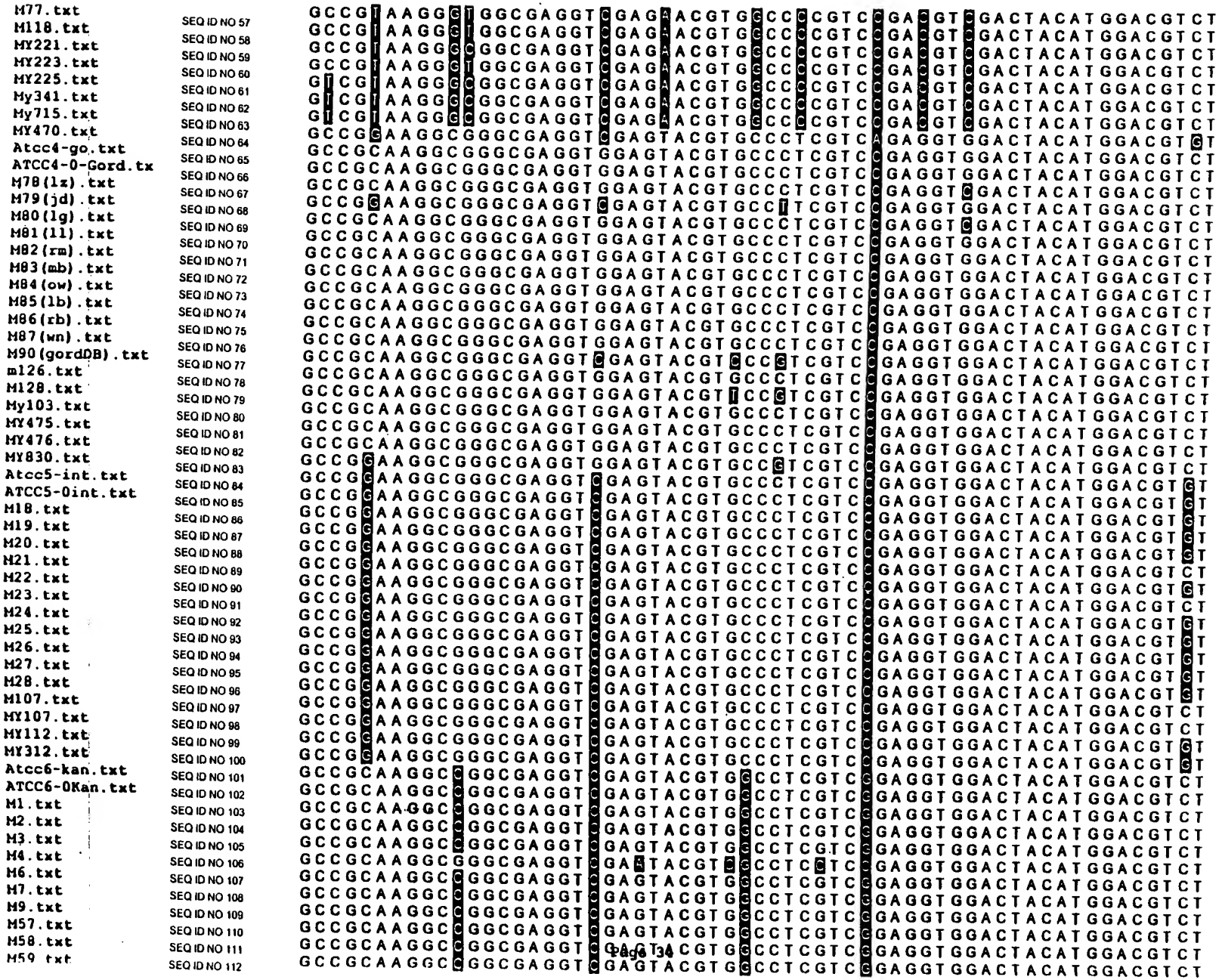

Table 1

BASE NOS 481-540

Table 1

BASE NOS 481-540

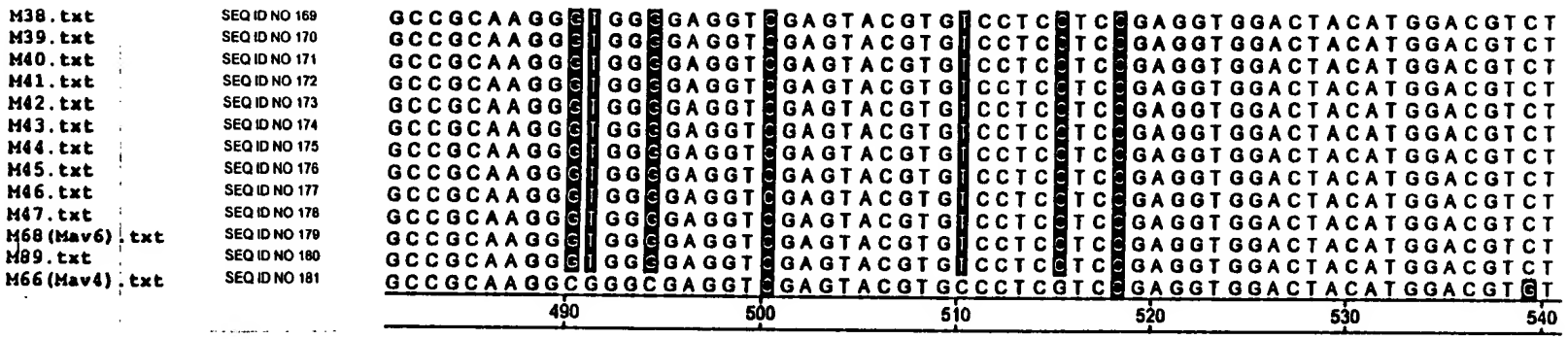

Table 1

BASE NOS 541-600

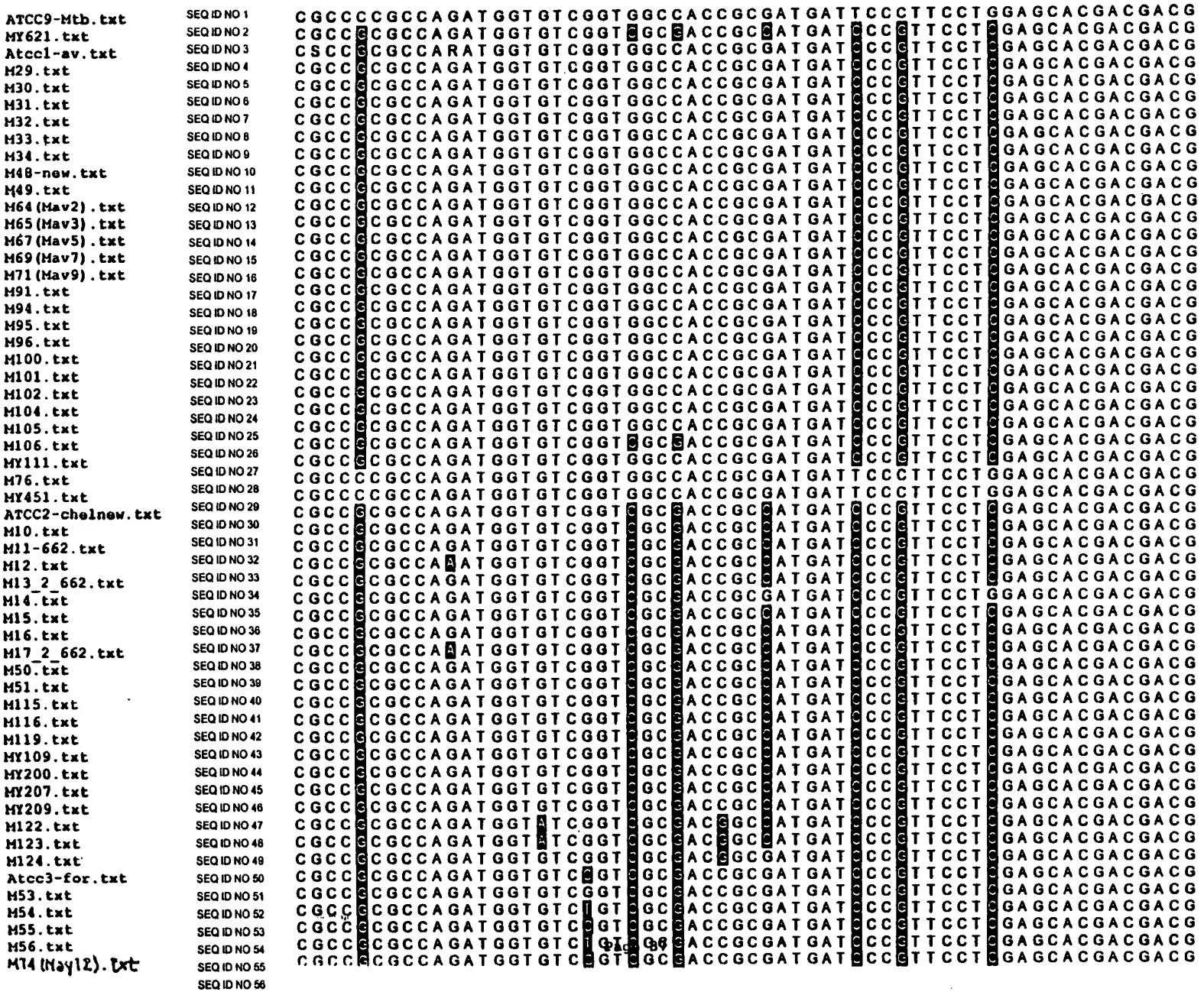

Table 1

BASE NOS 541-600

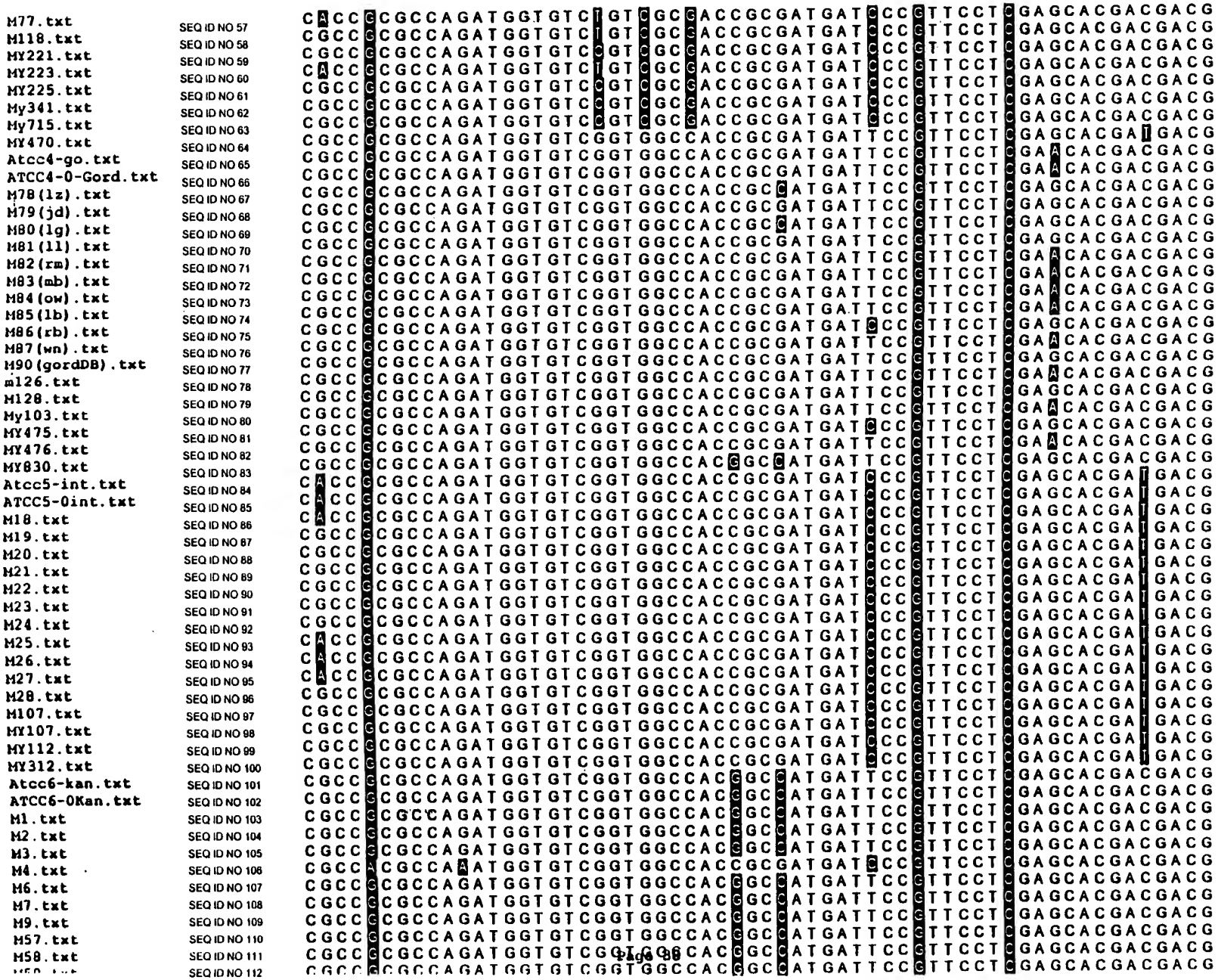

Table 1

BASE NOS 541-600

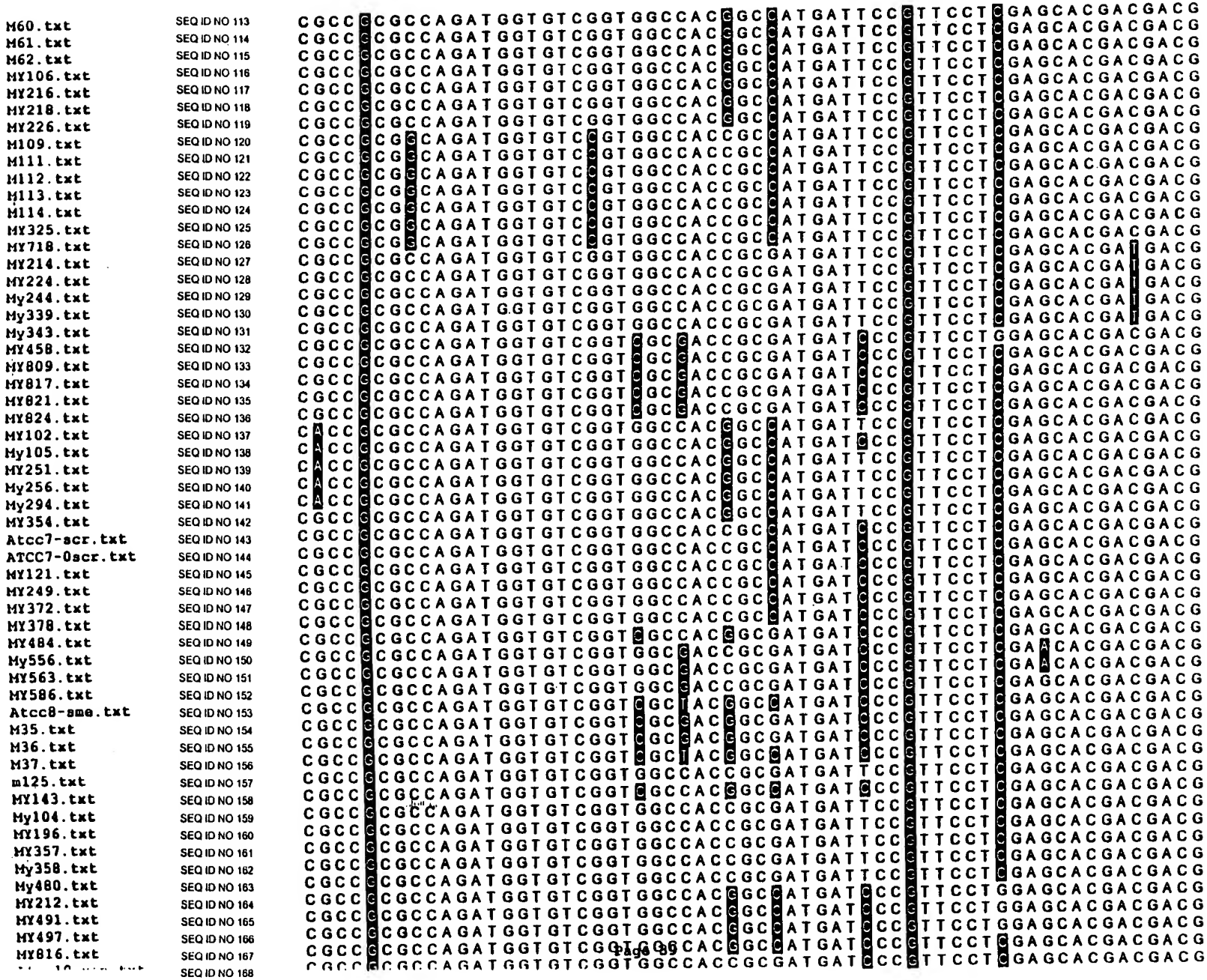

Table 1

BASE NOS 541-600

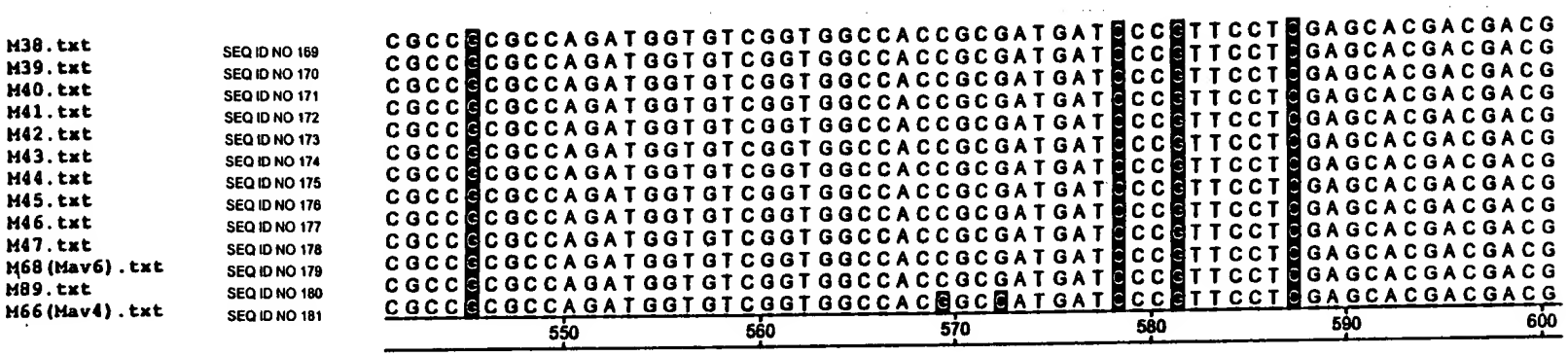

Table 1

BASE NOS 600-660

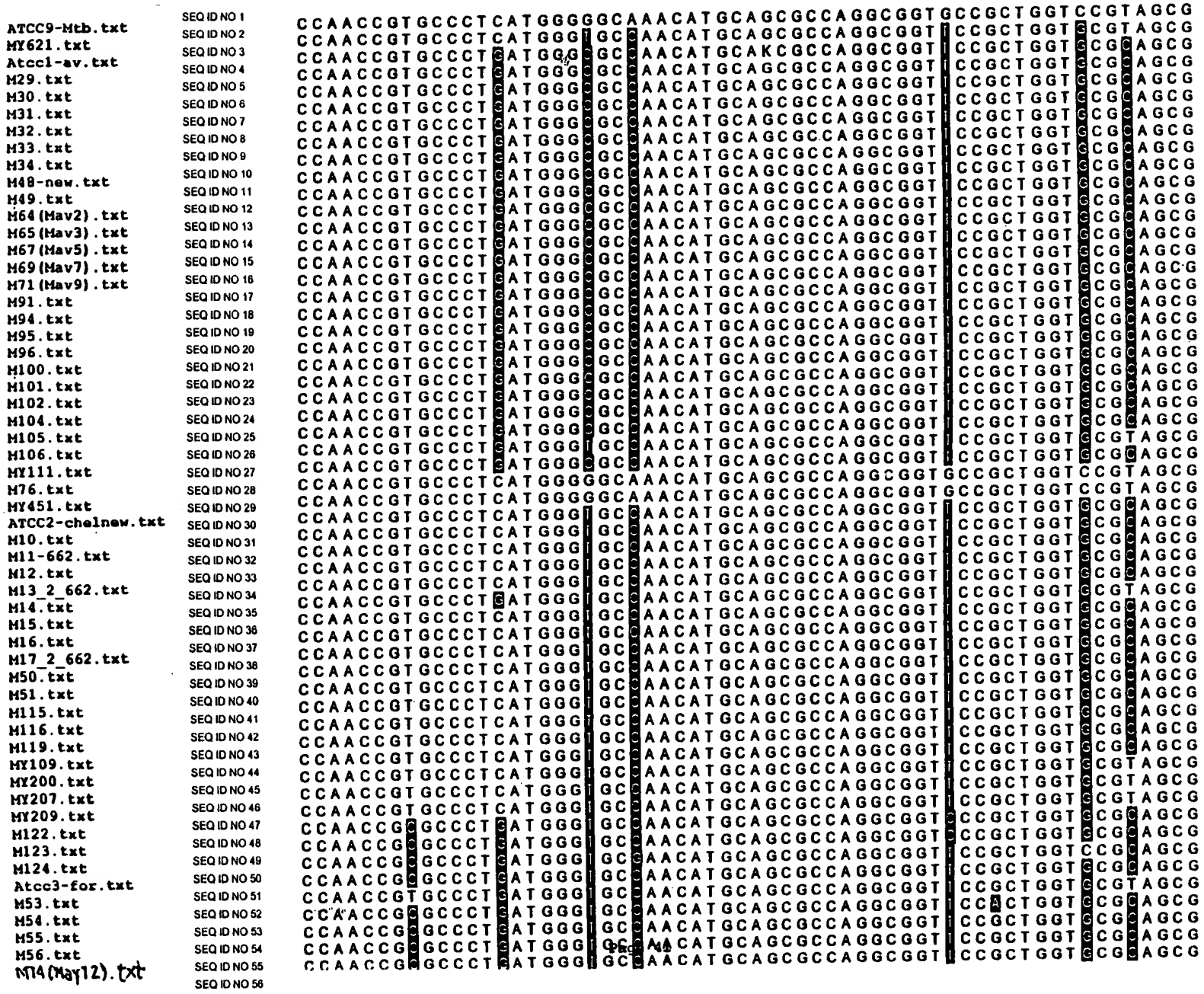

Table 1

BASE NOS 600-660

M77.txt	SEQ ID NO 57
M118.txt	SEQ ID NO 58
MY221.txt	SEQ ID NO 59
MY223.txt	SEQ ID NO 60
MY225.txt	SEQ ID NO 61
My341.txt	SEQ ID NO 62
My715.txt	SEQ ID NO 63
MY470.txt	SEQ ID NO 64
Atcc4-go.txt	SEQ ID NO 65
ATCC4-0-Gord.txt	SEQ ID NO 66
M78(lz).txt	SEQ ID NO 67
M79(jd).txt	SEQ ID NO 68
M80(lg).txt	SEQ ID NO 69
M81(ll).txt	SEQ ID NO 70
M82(rm).txt	SEQ ID NO 71
M83(mb).txt	SEQ ID NO 72
M84(ow).txt	SEQ ID NO 73
M85(lb).txt	SEQ ID NO 74
M86(rb).txt	SEQ ID NO 75
M87(wn).txt	SEQ ID NO 76
M90(gordDB).txt	SEQ ID NO 77
m126.txt	SEQ ID NO 78
M128.txt	SEQ ID NO 79
My103.txt	SEQ ID NO 80
MY475.txt	SEQ ID NO 81
MY476.txt	SEQ ID NO 82
MY830.txt	SEQ ID NO 83
Atcc5-int.txt	SEQ ID NO 84
ATCC5-0int.txt	SEQ ID NO 85
M18.txt	SEQ ID NO 86
M19.txt	SEQ ID NO 87
M20.txt	SEQ ID NO 88
M21.txt	SEQ ID NO 89
M22.txt	SEQ ID NO 90
M23.txt	SEQ ID NO 91
M24.txt	SEQ ID NO 92
M25.txt	SEQ ID NO 93
M26.txt	SEQ ID NO 94
M27.txt	SEQ ID NO 95
M28.txt	SEQ ID NO 96
M107.txt	SEQ ID NO 97
MY107.txt	SEQ ID NO 98
MY112.txt	SEQ ID NO 99
MY312.txt	SEQ ID NO 100
Atcc6-kan.txt	SEQ ID NO 101
ATCC6-0Kan.txt	SEQ ID NO 102
M1.txt	SEQ ID NO 103
M2.txt	SEQ ID NO 104
M3.txt	SEQ ID NO 105
M4.txt	SEQ ID NO 106
M6.txt	SEQ ID NO 107
M7.txt	SEQ ID NO 108
M9.txt	SEQ ID NO 109
M57.txt	SEQ ID NO 110
M58.txt	SEQ ID NO 111
M59.txt	SEQ ID NO 112

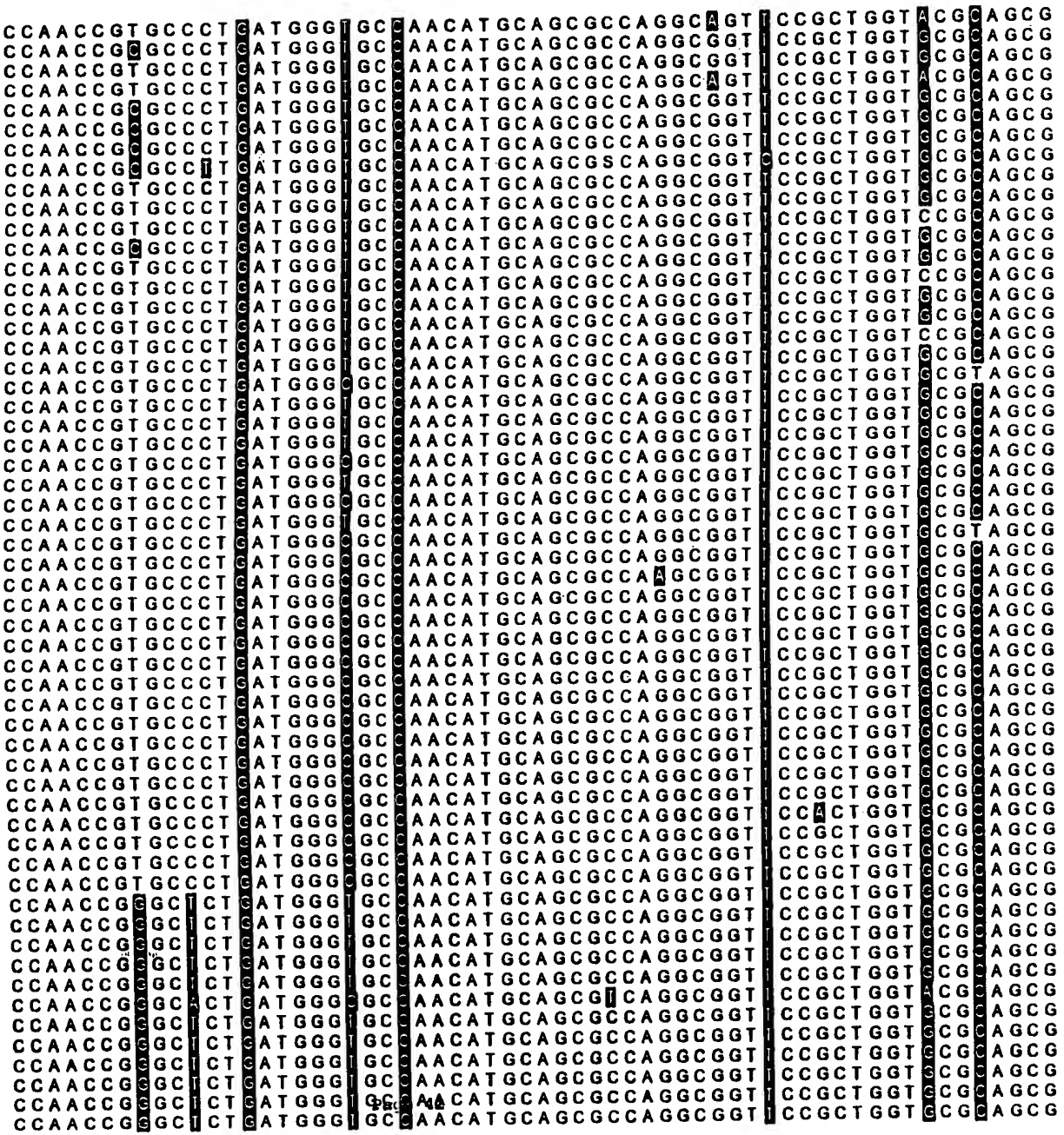

Table 1

BASE NOS 600-660

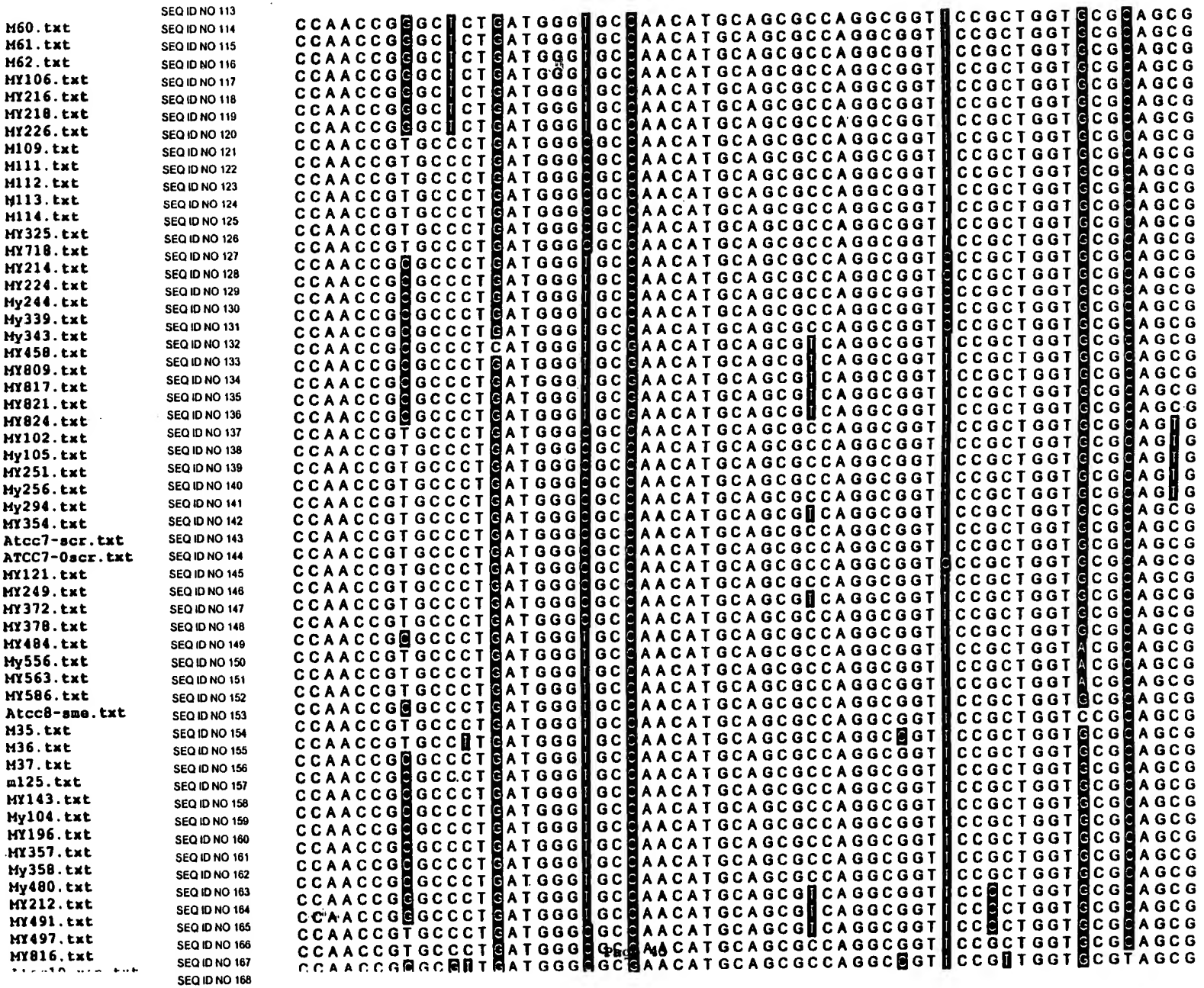

Table 1

BASE NOS 600-660

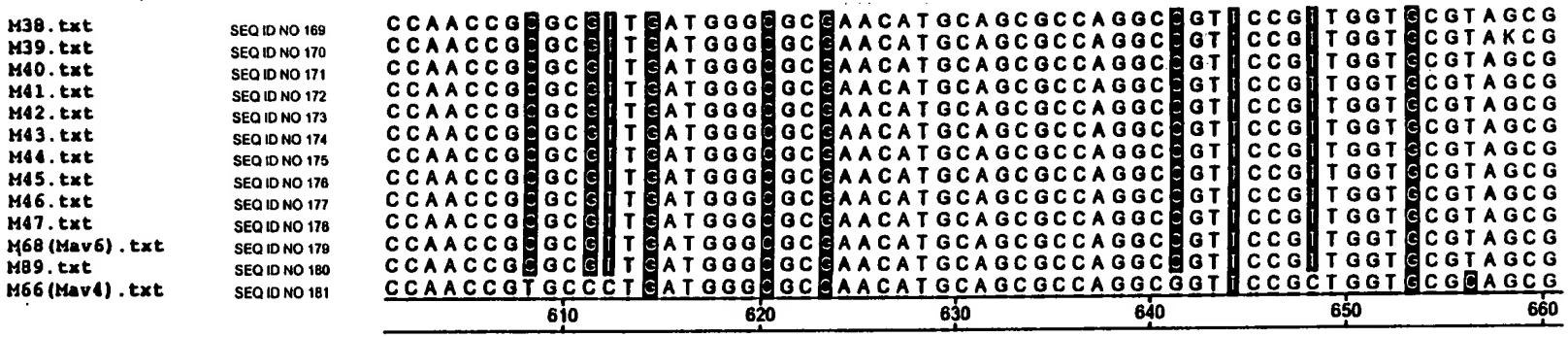

Table 1

BASE NOS 660-720

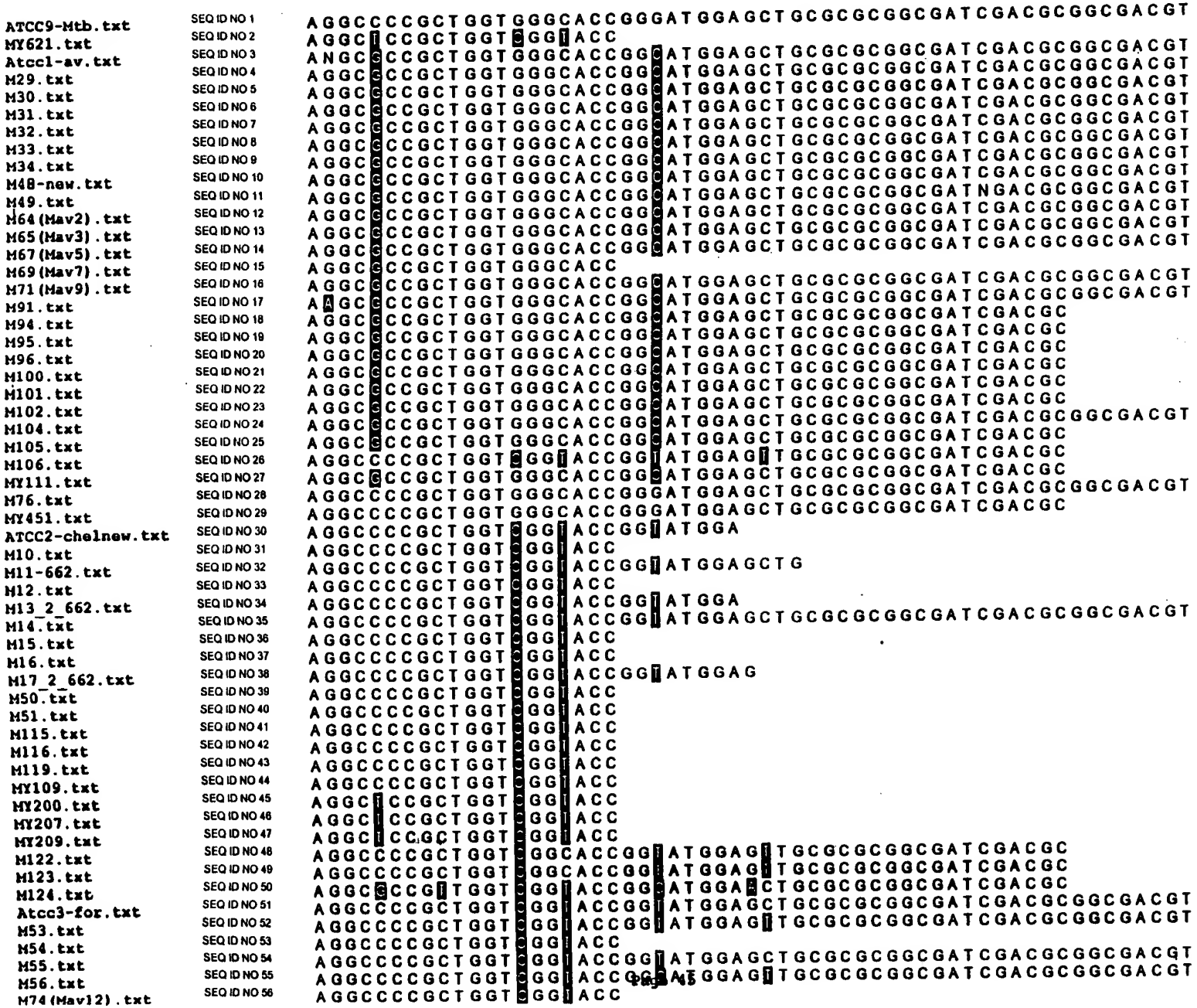

Table 1

BASE NOS 660-720

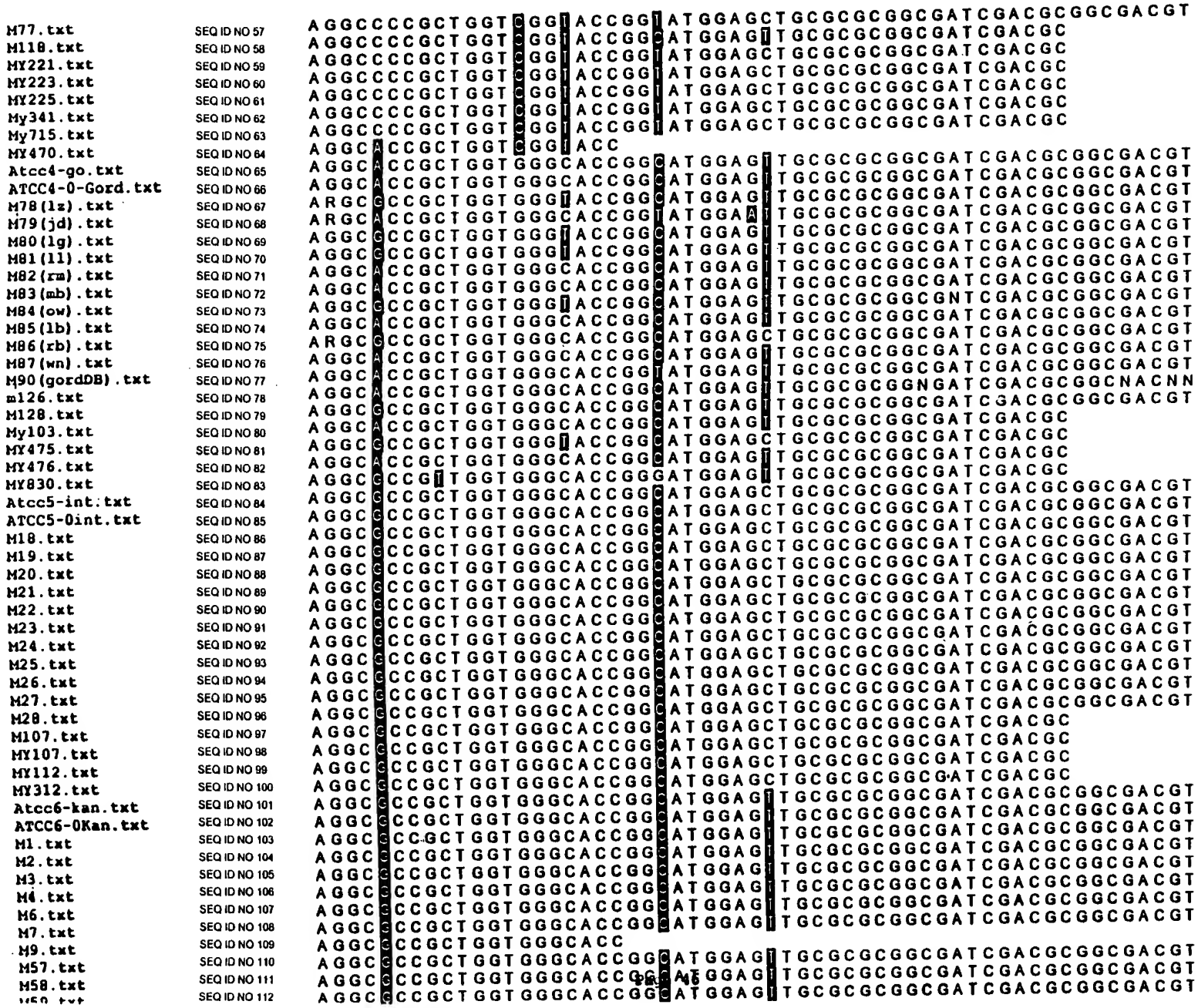

Table 1

BASE NOS 660-720

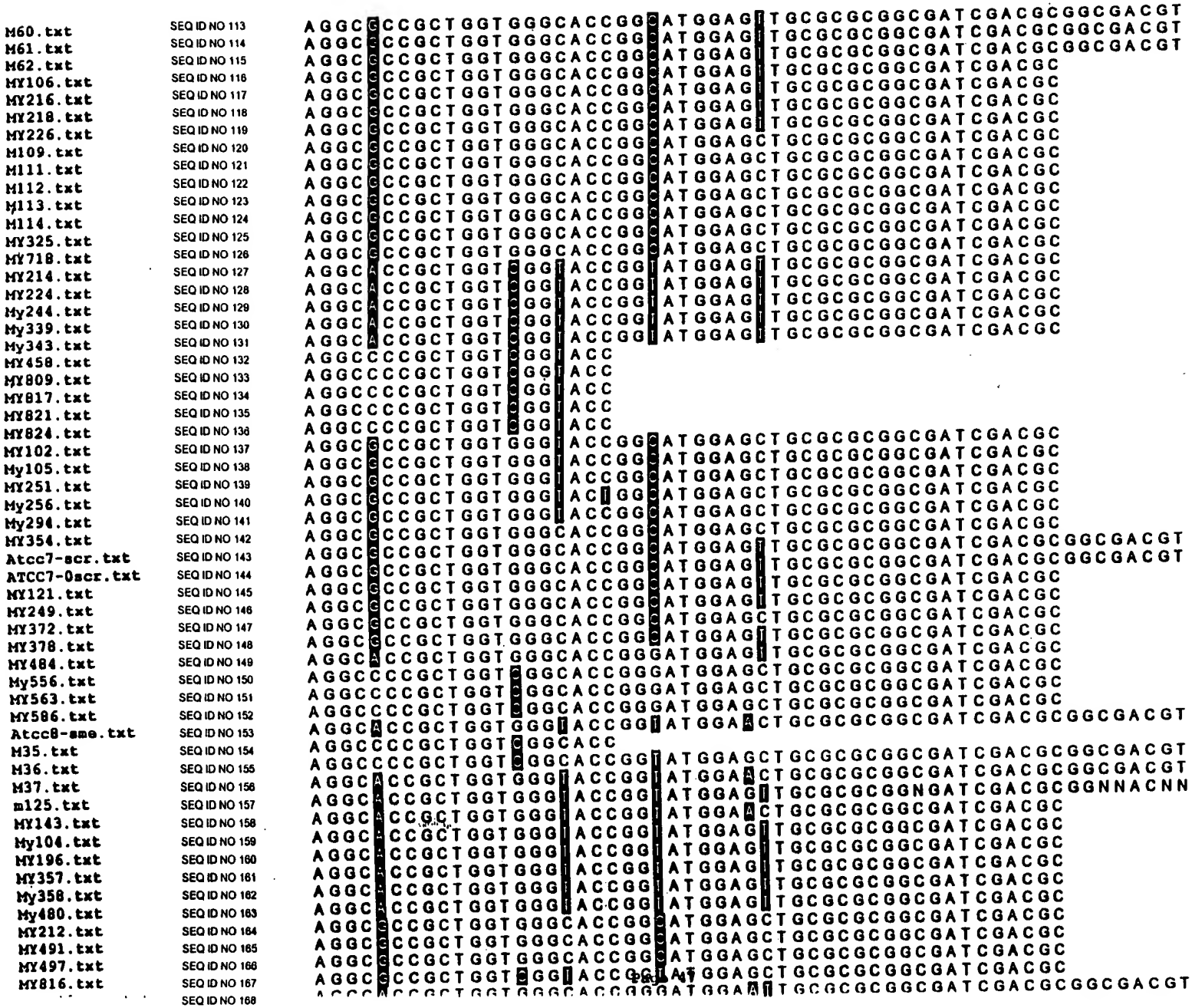

Table 1

BASE NOS 660-720

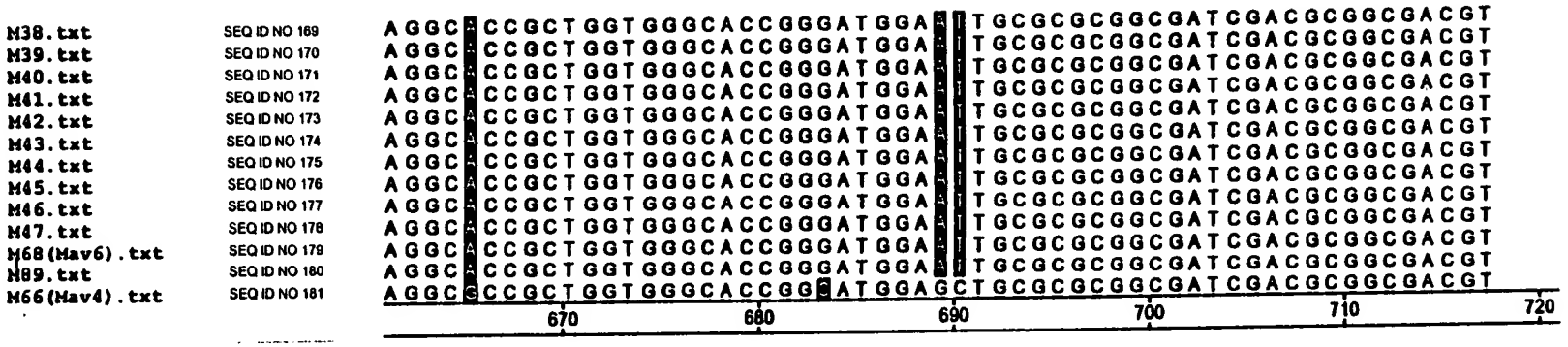

Table 2

Affy#	SAMPLE ID#	Alt. ID	SPECIES
MY621		ATCC	*M. abscessus*
ATCC1	25291	ATCC-av	*M. avium*
M100	60300	MAC	*M. avium*
M101	60112	MAC	*M. avium*
M102	60268	MAC	*M. avium*
M103	60270	MAC	*M. avium*
M104	60272	MAC	*M. avium*
M105	60293	MAC	*M. avium*
M106	60313	MAC	*M. avium*
M107	60345	MAC	*M. avium*
M29	95-1764		*M. avium*
M30	95-1766		*M. avium*
M31	95-1768		*M. avium*
M32	95-1770		*M. avium*
M33	95-1775		*M. avium*
M34	95-1776		*M. avium*
M48	95-1765		*M. avium*
M49	95-1769		*M. avium*
M83	MAC #1	MAC	*M. avium*
M84	MAC #2	MAC	*M. avium*
M85	MAC #3	MAC	*M. avium*
M87	MAC #5	MAC	*M. avium*
M89	MAC #7	MAC	*M. avium*
M70	MAC #8	MAC	*M. avium*
M71	MAC #9	MAC	*M. avium*
M72	MAC #10	MAC	*M. avium*
M91	FM	avium-intracell. FM(MAC)	*M. avium*
M92	60040	MAC	*M. avium*
M93	60042	MAC	*M. avium*
M94	60049	MAC	*M. avium*
M95	60051	MAC	*M. avium*
M96	60110	MAC	*M. avium*
M97	60116	MAC	*M. avium*
M98	60123	MAC	*M. avium*
M99	60176	MAC	*M. avium*
M76	92-773		*M. bovis*
MY451			*M. bovis*
ATCC2	35752	ATCC-chel(new)	*M. chelonae*
M10	95A9151		*M. chelonae*
M11	95A0477		*M. chelonae*
M115	60121		*M. chelonae*
M116	52942		*M. chelonae*
M117	43192		*M. chelonae*
M118	53160		*M. chelonae*
M119	53131		*M. chelonae*
M12	95A4883		*M. chelonae*
M120	52923		*M. chelonae*
M121	52919		*M. chelonae*
M13	95A2611		*M. chelonae*
M14	95A0779		*M. chelonae*
M15	95A8654		*M. chelonae*
M16	95A8882		*M. chelonae*
M17	95A8881		*M. chelonae*
M50	95A11814		*M. chelonae*
M51	95A1102		*M. chelonae*
M75	#13	MAC#13	*M. chelonae*
MY109			*M. chelonae*
MY200			*M. chelonae*

Table 2

MY207			*M. chelonae*
MY209			*M. chelonae*
M122	60025		*M. flavescens*
M123	60078		*M. flavescens*
M124	60252		*M. flavescens*
ATCC3	6841	ATCC-for	*M. fortuitum*
M53	60305		*M. fortuitum*
M54	60344		*M. fortuitum*
M55	60435		*M. fortuitum*
M56	60447		*M. fortuitum*
M74	#12	MAC#12	*M. fortuitum*
M88	CH	fort. CH	*M. fortuitum*
MY221			*M. fortuitum*
MY223			*M. fortuitum*
MY225			*M. fortuitum*
MY341			*M. fortuitum*
MY715			*M. fortuitum*
MY470			*M. genevese*
ATCC4	14470	ATCC-go	*M. gordonae*
ATCC4-0		gord.	*M. gordonae*
M125	60068		*M. gordonae*
M126	60182		*M. gordonae*
M127	60214		*M. gordonae*
M128	60283		*M. gordonae*
M78	92-842	gord. LZ	*M. gordonae*
M79	93-692	gord. JD	*M. gordonae*
M80	94-94	gord. LG	*M. gordonae*
M81	93-1231	gord. LL	*M. gordonae*
M82	93-463	gord. RM	*M. gordonae*
M83	92-1219	gord. MB	*M. gordonae*
M84	91-1131	gord. OW	*M. gordonae*
M85	91-1478	gord. LB	*M. gordonae*
M86	92-642	gord. RB	*M. gordonae*
M87	93-1180	gord. WN	*M. gordonae*
M90	DB	gord. DB	*M. gordonae*
MY103			*M. gordonae*
MY475			*M. gordonae*
MY476			*M. gordonae*
MY746			*M. gordonae*
MY830			*M. gordonae*
ATCC5		ATCC-int	*M. intracellulare*
ATCC5-0		intra	*M. intracellulare*
M18	95-1778		*M. intracellulare*
M19	95-1780		*M. intracellulare*
M20	95-1781		*M. intracellulare*
M21	95-1782		*M. intracellulare*
M22	95-1790		*M. intracellulare*
M23	95-1794		*M. intracellulare*
M24	95-1796		*M. intracellulare*
M25	95-1777		*M. intracellulare*
M26	95-1779		*M. intracellulare*
M27	95-1760		*M. intracellulare*
M28	95-1761		*M. intracellulare*
ATCC6	12478	ATCC-kan	*M. kansasii*
ATCC6-0		kans.	*M. kansasii*
M1	95A5375		*M. kansasii*
M2	95A10299		*M. kansasii*
M3	96A0020		*M. kansasii*
M4	95A3977		*M. kansasii*

Table 2

M5	95A4739		*M. kansasii*
M52	95A5381		*M. kansasii*
M57	60163		*M. kansasii*
M58	60180		*M. kansasii*
M59	60207		*M. kansasii*
M6	95A2695		*M. kansasii*
M60	60294		*M. kansasii*
M61	60308		*M. kansasii*
M62	60314		*M. kansasii*
M7	95A2694		*M. kansasii*
M73	#11	MAC#11	*M. kansasii*
M8	94A9042		*M. kansasii*
M9	95A1275		*M. kansasii*
MY106			*M. kansasii*
MY141			*M. kansasii*
MY216			*M. kansasii*
MY218			*M. kansasii*
MY226			*M. kansasii*
M108	60044		*M. malmoense*
M109	60149		*M. malmoense*
M110	60211		*M. malmoense*
M111	60202		*M. malmoense*
M112	60085		*M. malmoense*
M113	60047		*M. malmoense*
M114	60185		*M. malmoense*
MY325		ATCC	*M. malmoense*
MY718		malmo	*M. malmoense*
MY214			*M. marinum*
MY224			*M. marinum*
MY244			*M. marinum*
MY339			*M. marinum*
MY343			*M. marinum*
MY458		ATCC	*M. mucogenicum*
MY809			*M. mucogenicum*
MY817			*M. mucogenicum*
MY821			*M. mucogenicum*
MY824			*M. mucogenicum*
MY102			*M. nonchromagenicum*
MY105			*M. nonchromagenicum*
MY251			*M. nonchromagenicum*
MY256			*M. nonchromagenicum*
MY294			*M. nonchromagenicum*
ATCC7	19981	ATCC-scr	*M. scrofulaceum*
ATCC7-0		scrof.	*M. scrofulaceum*
MY121			*M. scrofulaceum*
MY249			*M. scrofulaceum*
MY372			*M. scrofulaceum*
MY378			*M. scrofulaceum*
MY484			*M. simiae*
MY556			*M. simiae*
MY563			*M. simiae*
MY586			*M. simiae*
ATCC8	19420	ATCC-sme	*M. smegmatis*
M35	95A1072		*M. smegmatis*
M36	95A8183		*M. smegmatis*
M37	95A4990		*M. smegmatis*
M77	92-144	smeg. JL	*M. smegmatis*
MY143		ATCC	*M. smegmatis*
MY104			*M. szulgai*

Table 2

MY196			*M. szulgai*
MY357			*M. szulgai*
MY358			*M. szulgai*
MY480			*M. szulgai*
TB74	C.17.96.5		*M. tab M160 DR*
MY387			*M. tb*
MY418			*M. tb*
MY437			*M. tb*
MY462			*M. tb*
TB59	C.18.96.1		*M. tb H37rv DR*
TB67	C.18.96.1		*M. tb H37rv DR*
TB73	C.17.96.1		*M. tb H37rv DR*
TB60	C.18.96.2		*M. tb J35 DR*
TB65	C.22.96.9		*M. tb M101 DR*
TB62	C.18.96.4		*M. tb M104 DR*
TB69	C.18.96.3		*M. tb M104 DR*
TB72	C.18.96.7		*M. tb M104DR*
TB66	C.22.96.10		*M. tb M112 DR*
TB63	C.18.96.5		*M. tb M140 DR*
TB64	C.18.96.6		*M. tb M160 DR*
TB70	C.18.96.4		*M. tb M160 DR*
TB61	C.18.96.3		*M. tb M60 DR*
TB68	C.18.96.2		*M. tb M60 DR*
TB71	C.18.96.6		*M. tb M60 DR*
MY212			*M. terrae*
MY354			*M. terrae*
MY491			*M. terrae*
MY497			*M. terrae*
MY816			*M. triplex*
ATCC9	27294	Mtb	*M. tuberculosis*
ATCC9-0		TB2020	*M. tuberculosis*
N/A	93-1071		*M. tuberculosis*
N/A	93-338		*M. tuberculosis*
N/A	92-852		*M. tuberculosis*
N/A	92-1005		*M. tuberculosis*
N/A	92-243		*M. tuberculosis*
N/A	92-304		*M. tuberculosis*
N/A	92-199		*M. tuberculosis*
N/A	92-197		*M. tuberculosis*
N/A	92-484		*M. tuberculosis*
N/A	94-577		*M. tuberculosis*
TB1	936		*M. tuberculosis*
TB10	1122		*M. tuberculosis*
TB11	3407		*M. tuberculosis*
TB12	978		*M. tuberculosis*
TB13	3553		*M. tuberculosis*
TB14	3466		*M. tuberculosis*
TB15	2163		*M. tuberculosis*
TB16	DW	DW	*M. tuberculosis*
TB17	CB	CB	*M. tuberculosis*
TB18	PB	PB	*M. tuberculosis*
TB19	AA	AA	*M. tuberculosis*
TB2	M0404A		*M. tuberculosis*
TB20	3492		*M. tuberculosis*
TB21	1435		*M. tuberculosis*
TB22	896		*M. tuberculosis*
TB23	2268		*M. tuberculosis*
TB24	3455		*M. tuberculosis*
TB25	37		*M. tuberculosis*

Table 2

TB26	173		*M. tuberculosis*
TB27	230		*M. tuberculosis*
TB28	2519		*M. tuberculosis*
TB29	T29233		*M. tuberculosis*
TB3	1231		*M. tuberculosis*
TB30	SP	SP	*M. tuberculosis*
TB31	3201		*M. tuberculosis*
TB32	3219		*M. tuberculosis*
TB33	80		*M. tuberculosis*
TB34	3442		*M. tuberculosis*
TB35	3502		*M. tuberculosis*
TB36	3759		*M. tuberculosis*
TB37	1295		*M. tuberculosis*
TB38	337		*M. tuberculosis*
TB39	394		*M. tuberculosis*
TB4	914		*M. tuberculosis*
TB40	499		*M. tuberculosis*
TB41	535		*M. tuberculosis*
TB42	607		*M. tuberculosis*
TB43	707		*M. tuberculosis*
TB44	692		*M. tuberculosis*
TB45	2408		*M. tuberculosis*
TB46	1069		*M. tuberculosis*
TB47	M3282A		*M. tuberculosis*
TB48	1338		*M. tuberculosis*
TB49	1388		*M. tuberculosis*
TB5	1145		*M. tuberculosis*
TB50	65		*M. tuberculosis*
TB51	727		*M. tuberculosis*
TB52	3455		*M. tuberculosis*
TB53	3508		*M. tuberculosis*
TB54	9600367		*M. tuberculosis*
TB55	9600173		*M. tuberculosis*
TB56	9503471		*M. tuberculosis*
TB57	9600309		*M. tuberculosis*
TB58	9600230		*M. tuberculosis*
TB6	1417		*M. tuberculosis*
TB7	SM2341		*M. tuberculosis*
TB75	2098		*M. tuberculosis*
TB76	173/1		*M. tuberculosis*
TB77	1122/1		*M. tuberculosis*
TB78	1417/1		*M. tuberculosis*
TB8	1587		*M. tuberculosis*
TB9	M7032A		*M. tuberculosis*
ATCC10	19250	*ATCC-xen*	*M. xenopi*
M129	60133		*M. xenopi*
M130	60200		*M. xenopi*
M131	60365		*M. xenopi*
M132	60367		*M. xenopi*
M38	95A5208		*M. xenopi*
M39	95A5399		*M. xenopi*
M40	95A3938		*M. xenopi*
M41	95A6782		*M. xenopi*
M42	95A0933		*M. xenopi*
M43	95A4320		*M. xenopi*
M44	95A3478		*M. xenopi*
M45	95A2997		*M. xenopi*
M46	95A8383		*M. xenopi*
M47	95A4319		*M. xenopi*

Table 2

M68	MAC #6	*MAC #6*	*M. xenopi*
M89	SG	xen. SG	*M. xenopi*
MY219			*M. xenopi*
MY250			*M. xenopi*
MY252			*M. xenopi*
MY254			*M. xenopi*
MY255			*M. xenopi*
MY107			*MAC*
MY111			*MAC*
MY112			*MAC*
MY312			*MAC*
M66	MAC #4	*MAC #4*	*unique*